



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/31, C07K 14/22, A61K 39/095, G01N 33/53, C12Q 1/68, C07K 16/12		A2	(11) International Publication Number: WO 99/36544
			(43) International Publication Date: 22 July 1999 (22.07.99)
(21) International Application Number: PCT/IB99/00103		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
(22) International Filing Date: 14 January 1999 (14.01.99)		Published <i>Without international search report and to be republished upon receipt of that report.</i>	
(30) Priority Data:			
9800760.2 14 January 1998 (14.01.98) GB			
9819015.0 1 September 1998 (01.09.98) GB			
9822143.5 9 October 1998 (09.10.98) GB			
(71) Applicant (for all designated States except US): CHIRON S.P.A. [IT/IT]; Via Fiorentina, 1, I-53100 Siena (IT).			
(72) Inventors; and			
(75) Inventors/Applicants (for US only): MASIGNANI, Vega [IT/IT]; Via Pantaneto, 105, I-53100 Siena (IT). RAPPUOLI, Rino [IT/IT]; Via delle Rocche, 1, Vagliagli, I-53019 Castelnuovo Berardenga (IT). PIZZA, Mariagrazia [IT/IT]; Strada di Montalbuccio, 160, I-53100 Siena (IT). SCARLATO, Vincenzo [IT/IT]; Via Firenze, 3/37, I-53134 Colle Val d'Elsa (IT). GRANDI, Guido [IT/IT]; 9° Strada, 4, I-20090 Segrate (IT).			
(74) Agent: HALLYBONE, Huw, George; Carpmals & Ransford, 43 Bloomsbury Square, London WC1A 2RA (GB).			
(54) Title: NEISSERIA MENINGITIDIS ANTIGENS			
(57) Abstract			
<p>The invention provides proteins from <i>Neisseria meningitidis</i> (strains A and B), including amino acid sequences, the corresponding nucleotide sequences, expression data, and serological data. The proteins are useful antigens for vaccines, immunogenic compositions, and/or diagnostics.</p>			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

NEISSERIA MENINGITIDIS ANTIGENS

This invention relates to antigens from the bacterium *Neisseria meningitidis*.

BACKGROUND

Neisseria meningitidis is a non-motile, gram negative diplococcus human pathogen. It colonises the pharynx, causing meningitis and, occasionally, septicaemia in the absence of meningitis. It is closely related to *N.gonorrhoeae*, although one feature that clearly differentiates meningococcus from gonococcus is the presence of a polysaccharide capsule that is present in all pathogenic meningococci.

N.meningitidis causes both endemic and epidemic disease. In the United States the attack rate is 0.6-1 per 100,000 persons per year, and it can be much greater during outbreaks (see Lieberman *et al.* (1996) Safety and Immunogenicity of a Serogroups A/C *Neisseria meningitidis* Oligosaccharide-Protein Conjugate Vaccine in Young Children. *JAMA* 275(19):1499-1503; Schuchat *et al* (1997) Bacterial Meningitis in the United States in 1995. *N Engl J Med* 337(14):970-976). In developing countries, endemic disease rates are much higher and during epidemics incidence rates can reach 500 cases per 100,000 persons per year. Mortality is extremely high, at 10-20% in the United States, and much higher in developing countries. Following the introduction of the conjugate vaccine against *Haemophilus influenzae*, *N. meningitidis* is the major cause of bacterial meningitis at all ages in the United States (Schuchat *et al* (1997) *supra*).

Based on the organism's capsular polysaccharide, 12 serogroups of *N.meningitidis* have been identified. Group A is the pathogen most often implicated in epidemic disease in sub-Saharan Africa. Serogroups B and C are responsible for the vast majority of cases in the United States and in most developed countries. Serogroups W135 and Y are responsible for the rest of the cases in the United States and developed countries. The meningococcal vaccine currently in use is a tetravalent polysaccharide vaccine composed of serogroups A, C, Y and W135. Although efficacious in adolescents and adults, it induces a poor immune response and short duration of protection, and cannot be used in infants [eg. Morbidity and Mortality weekly report, Vol.46, No. RR-5 (1997)]. This is because polysaccharides are T-cell independent antigens that induce a weak immune response that cannot be boosted by repeated immunization. Following the success of the

vaccination against *H. influenzae*, conjugate vaccines against serogroups A and C have been developed and are at the final stage of clinical testing (Zollinger WD "New and Improved Vaccines Against Meningococcal Disease" in: *New Generation Vaccines, supra*, pp. 469-488; Lieberman *et al* (1996) *supra*; Costantino *et al* (1992) Development and phase I clinical testing of a conjugate vaccine against meningococcus A and C. *Vaccine* 10:691-698).

Meningococcus B remains a problem, however. This serotype currently is responsible for approximately 50% of total meningitis in the United States, Europe, and South America. The polysaccharide approach cannot be used because the menB capsular polysaccharide is a polymer of $\alpha(2-8)$ -linked *N*-acetyl neuraminic acid that is also present in mammalian tissue. This results in tolerance to the antigen; indeed, if an immune response were elicited, it would be anti-self, and therefore undesirable. In order to avoid induction of autoimmunity and to induce a protective immune response, the capsular polysaccharide has, for instance, been chemically modified substituting the *N*-acetyl groups with *N*-propionyl groups, leaving the specific antigenicity unaltered (Romero & Outschoorn (1994) Current status of Meningococcal group B vaccine candidates: capsular or non-capsular? *Clin Microbiol Rev* 7(4):559-575).

Alternative approaches to menB vaccines have used complex mixtures of outer membrane proteins (OMPs), containing either the OMPs alone, or OMPs enriched in porins, or deleted of the class 4 OMPs that are believed to induce antibodies that block bactericidal activity. This approach produces vaccines that are not well characterized. They are able to protect against the homologous strain, but are not effective at large where there are many antigenic variants of the outer membrane proteins. To overcome the antigenic variability, multivalent vaccines containing up to nine different porins have been constructed (*eg.* Poolman JT (1992) Development of a meningococcal vaccine. *Infect. Agents Dis.* 4:13-28). Additional proteins to be used in outer membrane vaccines have been the opa and opc proteins, but none of these approaches have been able to overcome the antigenic variability (*eg.* Ala'Aldeen & Borriello (1996) The meningococcal transferrin-binding proteins 1 and 2 are both surface exposed and generate bactericidal antibodies capable of killing homologous and heterologous strains. *Vaccine* 14(1):49-53).

A certain amount of sequence data is available for meningococcal and gonococcal genes and proteins (*eg.* EP-A-0467714, WO96/29412), but this is by no means complete. The provision of further sequences could provide an opportunity to identify secreted or surface-exposed proteins that

are presumed targets for the immune system and which are not antigenically variable. For instance, some of the identified proteins could be components of efficacious vaccines against meningococcus B, some could be components of vaccines against all meningococcal serotypes, and others could be components of vaccines against all pathogenic *Neisseriae*.

5 THE INVENTION

The invention provides proteins comprising the *N.meningitidis* amino acid sequences disclosed in the examples.

It also provides proteins comprising sequences homologous (*ie.* having sequence identity) to the *N.meningitidis* amino acid sequences disclosed in the examples. Depending on the particular
10 sequence, the degree of sequence identity is preferably greater than 50% (*eg.* 60%, 70%, 80%, 90%, 95%, 99% or more). These homologous proteins include mutants and allelic variants of the sequences disclosed in the examples. Typically, 50% identity or more between two proteins is considered to be an indication of functional equivalence. Identity between the proteins is preferably determined by the Smith-Waterman homology search algorithm as implemented in the MPSRCH
15 program (Oxford Molecular), using an affine gap search with parameters *gap open penalty=12* and *gap extension penalty=1*.

The invention further provides proteins comprising fragments of the *N.meningitidis* amino acid sequences disclosed in the examples. The fragments should comprise at least *n* consecutive amino acids from the sequences and, depending on the particular sequence, *n* is 7 or more (*eg.* 8, 10, 12,
20 14, 16, 18, 20 or more). Preferably the fragments comprise an epitope from the sequence.

The proteins of the invention can, of course, be prepared by various means (*eg.* recombinant expression, purification from cell culture, chemical synthesis *etc.*) and in various forms (*eg.* native, fusions *etc.*). They are preferably prepared in substantially pure form (*ie.* substantially free from other *N.meningitidis* or host cell proteins)

25 According to a further aspect, the invention provides antibodies which bind to these proteins. These may be polyclonal or monoclonal and may be produced by any suitable means.

According to a further aspect, the invention provides nucleic acid comprising the *N.meningitidis* nucleotide sequences disclosed in the examples. In addition, the invention provides nucleic acid comprising sequences homologous (*ie.* having sequence identity) to the *N.meningitidis* nucleotide sequences disclosed in the examples.

- 5 Furthermore, the invention provides nucleic acid which can hybridise to the *N.meningitidis* nucleic acid disclosed in the examples, preferably under "high stringency" conditions (*eg.* 65°C in a 0.1xSSC, 0.5% SDS solution).

- Nucleic acid comprising fragments of these sequences are also provided. These should comprise at least *n* consecutive nucleotides from the *N.meningitidis* sequences and, depending on the
10 particular sequence, *n* is 10 or more (*eg.* 12, 14, 15, 18, 20, 25, 30, 35, 40 or more).

According to a further aspect, the invention provides nucleic acid encoding the proteins and protein fragments of the invention.

It should also be appreciated that the invention provides nucleic acid comprising sequences complementary to those described above (*eg.* for antisense or probing purposes).

- 15 Nucleic acid according to the invention can, of course, be prepared in many ways (*eg.* by chemical synthesis, from genomic or cDNA libraries, from the organism itself *etc.*) and can take various forms (*eg.* single stranded, double stranded, vectors, probes *etc.*).

In addition, the term "nucleic acid" includes DNA and RNA, and also their analogues, such as those containing modified backbones, and also peptide nucleic acids (PNA) *etc.*

- 20 According to a further aspect, the invention provides vectors comprising nucleotide sequences of the invention (*eg.* expression vectors) and host cells transformed with such vectors.

According to a further aspect, the invention provides compositions comprising protein, antibody, and/or nucleic acid according to the invention. These compositions may be suitable as vaccines, for instance, or as diagnostic reagents, or as immunogenic compositions.

The invention also provides nucleic acid, protein, or antibody according to the invention for use as medicaments (eg. as vaccines) or as diagnostic reagents. It also provides the use of nucleic acid, protein, or antibody according to the invention in the manufacture of: (i) a medicament for treating or preventing infection due to Neisserial bacteria; (ii) a diagnostic reagent for detecting the presence of Neisserial bacteria or of antibodies raised against Neisserial bacteria; and/or (iii) a reagent which can raise antibodies against Neisserial bacteria. Said Neisserial bacteria may be any species or strain (such as *N.gonorrhoeae*) but are preferably *N.meningitidis*, especially strain A, strain B or strain C.

The invention also provides a method of treating a patient, comprising administering to the patient a therapeutically effective amount of nucleic acid, protein, and/or antibody according to the invention.

According to further aspects, the invention provides various processes.

A process for producing proteins of the invention is provided, comprising the step of culturing a host cell according to the invention under conditions which induce protein expression.

A process for producing protein or nucleic acid of the invention is provided, wherein the protein or nucleic acid is synthesised in part or in whole using chemical means.

A process for detecting polynucleotides of the invention is provided, comprising the steps of: (a) contacting a nucleic probe according to the invention with a biological sample under hybridizing conditions to form duplexes; and (b) detecting said duplexes.

A process for detecting proteins of the invention is provided, comprising the steps of: (a) contacting an antibody according to the invention with a biological sample under conditions suitable for the formation of an antibody-antigen complexes; and (b) detecting said complexes.

Unlike the sequences disclosed in PCT/IB98/01665, the sequences disclosed in the present application are believed not to have any significant homologs in *N.gonorrhoeae*. Accordingly, the sequences of the present invention also find use in the preparation of reagents for distinguishing between *N.meningitidis* and *N.gonorrhoeae*

A summary of standard techniques and procedures which may be employed in order to perform the invention (eg. to utilise the disclosed sequences for vaccination or diagnostic purposes) follows. This summary is not a limitation on the invention but, rather, gives examples that may be used, but are not required.

5 General

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature eg. Sambrook *Molecular Cloning; A Laboratory Manual, Second Edition* (1989); *DNA Cloning, Volumes I and ii* (D.N Glover ed. 1985); *Oligonucleotide Synthesis* (M.J. Gait ed, 1984); *Nucleic Acid Hybridization* (B.D. Hames & S.J. Higgins eds. 1984); *Transcription and Translation* (B.D. Hames & S.J. Higgins eds. 1984); *Animal Cell Culture* (R.I. Freshney ed. 1986); *Immobilized Cells and Enzymes* (IRL Press, 1986); B. Perbal, *A Practical Guide to Molecular Cloning* (1984); the *Methods in Enzymology* series (Academic Press, Inc.), especially volumes 154 & 155; *Gene Transfer Vectors for Mammalian Cells* (J.H. Miller and M.P. Calos eds. 1987, Cold Spring Harbor Laboratory); Mayer and Walker, eds. (1987), *Immunochemical Methods in Cell and Molecular Biology* (Academic Press, London); Scopes, (1987) *Protein Purification: Principles and Practice*, Second Edition (Springer-Verlag, N.Y.), and *Handbook of Experimental Immunology, Volumes I-IV* (D.M. Weir and C. C. Blackwell eds 1986).

20 Standard abbreviations for nucleotides and amino acids are used in this specification.

All publications, patents, and patent applications cited herein are incorporated in full by reference. In particular, the contents of UK patent applications 9800760.2, 9819015.0 and 9822143.5 are incorporated herein.

Definitions

25 A composition containing X is "substantially free of" Y when at least 85% by weight of the total X+Y in the composition is X. Preferably, X comprises at least about 90% by weight of the total of X+Y in the composition, more preferably at least about 95% or even 99% by weight.

The term "comprising" means "including" as well as "consisting" *eg.* a composition "comprising" X may consist exclusively of X or may include something additional to X, such as X+Y.

The term "heterologous" refers to two biological components that are not found together in nature.

The components may be host cells, genes, or regulatory regions, such as promoters. Although the

5 heterologous components are not found together in nature, they can function together, as when a promoter heterologous to a gene is operably linked to the gene. Another example is where a Neisserial sequence is heterologous to a mouse host cell. A further examples would be two epitopes from the same or different proteins which have been assembled in a single protein in an arrangement not found in nature.

10 An "origin of replication" is a polynucleotide sequence that initiates and regulates replication of polynucleotides, such as an expression vector. The origin of replication behaves as an autonomous unit of polynucleotide replication within a cell, capable of replication under its own control. An origin of replication may be needed for a vector to replicate in a particular host cell. With certain origins of replication, an expression vector can be reproduced at a high copy number in the
15 presence of the appropriate proteins within the cell. Examples of origins are the autonomously replicating sequences, which are effective in yeast; and the viral T-antigen, effective in COS-7 cells.

A "mutant" sequence is defined as DNA, RNA or amino acid sequence differing from but having sequence identity with the native or disclosed sequence. Depending on the particular sequence, the
20 degree of sequence identity between the native or disclosed sequence and the mutant sequence is preferably greater than 50% (*eg.* 60%, 70%, 80%, 90%, 95%, 99% or more, calculated using the Smith-Waterman algorithm as described above). As used herein, an "allelic variant" of a nucleic acid molecule, or region, for which nucleic acid sequence is provided herein is a nucleic acid molecule, or region, that occurs essentially at the same locus in the genome of another or second
25 isolate, and that, due to natural variation caused by, for example, mutation or recombination, has a similar but not identical nucleic acid sequence. A coding region allelic variant typically encodes a protein having similar activity to that of the protein encoded by the gene to which it is being compared. An allelic variant can also comprise an alteration in the 5' or 3' untranslated regions of the gene, such as in regulatory control regions (*eg.* see US patent 5,753,235).

Expression systems

The Neisserial nucleotide sequences can be expressed in a variety of different expression systems; for example those used with mammalian cells, baculoviruses, plants, bacteria, and yeast.

i. Mammalian Systems

- 5 Mammalian expression systems are known in the art. A mammalian promoter is any DNA sequence capable of binding mammalian RNA polymerase and initiating the downstream (3') transcription of a coding sequence (eg. structural gene) into mRNA. A promoter will have a transcription initiating region, which is usually placed proximal to the 5' end of the coding sequence, and a TATA box, usually located 25-30 base pairs (bp) upstream of the transcription
10 initiation site. The TATA box is thought to direct RNA polymerase II to begin RNA synthesis at the correct site. A mammalian promoter will also contain an upstream promoter element, usually located within 100 to 200 bp upstream of the TATA box. An upstream promoter element determines the rate at which transcription is initiated and can act in either orientation [Sambrook et al. (1989) "Expression of Cloned Genes in Mammalian Cells." In *Molecular Cloning: A*
15 *Laboratory Manual*, 2nd ed.]

Mammalian viral genes are often highly expressed and have a broad host range; therefore sequences encoding mammalian viral genes provide particularly useful promoter sequences. Examples include the SV40 early promoter, mouse mammary tumor virus LTR promoter, adenovirus major late promoter (Ad MLP), and herpes simplex virus promoter. In addition, sequences derived from non-
20 viral genes, such as the murine metallothionein gene, also provide useful promoter sequences. Expression may be either constitutive or regulated (inducible), depending on the promoter can be induced with glucocorticoid in hormone-responsive cells.

The presence of an enhancer element (enhancer), combined with the promoter elements described above, will usually increase expression levels. An enhancer is a regulatory DNA sequence that can
25 stimulate transcription up to 1000-fold when linked to homologous or heterologous promoters, with synthesis beginning at the normal RNA start site. Enhancers are also active when they are placed upstream or downstream from the transcription initiation site, in either normal or flipped orientation, or at a distance of more than 1000 nucleotides from the promoter [Maniatis et al. (1987) *Science* 236:1237; Alberts et al. (1989) *Molecular Biology of the Cell*, 2nd ed.]. Enhancer elements
30 derived from viruses may be particularly useful, because they usually have a broader host range.

Examples include the SV40 early gene enhancer [Dijkema et al (1985) *EMBO J.* 4:761] and the enhancer/promoters derived from the long terminal repeat (LTR) of the Rous Sarcoma Virus [Gorman et al. (1982b) *Proc. Natl. Acad. Sci.* 79:6777] and from human cytomegalovirus [Boshart et al. (1985) *Cell* 41:521]. Additionally, some enhancers are regulatable and become active only
5 in the presence of an inducer, such as a hormone or metal ion [Sassone-Corsi and Borelli (1986) *Trends Genet.* 2:215; Maniatis et al. (1987) *Science* 236:1237].

A DNA molecule may be expressed intracellularly in mammalian cells. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always be a methionine, which is encoded by the ATG start codon. If desired,
10 the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide.

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provides for secretion of the foreign protein in mammalian cells. Preferably, there are processing sites encoded between the leader fragment and the foreign gene that can be cleaved either *in vivo*
15 or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The adenovirus tripartite leader is an example of a leader sequence that provides for secretion of a foreign protein in mammalian cells.

Usually, transcription termination and polyadenylation sequences recognized by mammalian cells
20 are regulatory regions located 3' to the translation stop codon and thus, together with the promoter elements, flank the coding sequence. The 3' terminus of the mature mRNA is formed by site-specific post-transcriptional cleavage and polyadenylation [Birnstiel et al. (1985) *Cell* 41:349; Proudfoot and Whitelaw (1988) "Termination and 3' end processing of eukaryotic RNA. In *Transcription and splicing* (ed. B.D. Hames and D.M. Glover); Proudfoot (1989) *Trends Biochem.*
25 *Sci.* 14:105]. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator/polyadenylation signals include those derived from SV40 [Sambrook et al (1989) "Expression of cloned genes in cultured mammalian cells." In *Molecular Cloning: A Laboratory Manual*].

Usually, the above described components, comprising a promoter, polyadenylation signal, and transcription termination sequence are put together into expression constructs. Enhancers, introns with functional splice donor and acceptor sites, and leader sequences may also be included in an expression construct, if desired. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (eg. plasmids) capable of stable maintenance in a host, such as mammalian cells or bacteria. Mammalian replication systems include those derived from animal viruses, which require trans-acting factors to replicate. For example, plasmids containing the replication systems of papovaviruses, such as SV40 [Gluzman (1981) *Cell* 23:175] or polyomavirus, replicate to extremely high copy number in the presence of the appropriate viral T antigen. Additional examples of mammalian replicons include those derived from bovine papillomavirus and Epstein-Barr virus. Additionally, the replicon may have two replicaton systems, thus allowing it to be maintained, for example, in mammalian cells for expression and in a prokaryotic host for cloning and amplification. Examples of such mammalian-bacteria shuttle vectors include pMT2 [Kaufman et al. (1989) *Mol. Cell. Biol.* 9:946] and pHEBO [Shimizu et al. (1986) *Mol. Cell. Biol.* 6:1074].

The transformation procedure used depends upon the host to be transformed. Methods for introduction of heterologous polynucleotides into mammalian cells are known in the art and include dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei.

Mammalian cell lines available as hosts for expression are known in the art and include many immortalized cell lines available from the American Type Culture Collection (ATCC), including but not limited to, Chinese hamster ovary (CHO) cells, HeLa cells, baby hamster kidney (BHK) cells, monkey kidney cells (COS), human hepatocellular carcinoma cells (eg. Hep G2), and a number of other cell lines.

ii. Baculovirus Systems

The polynucleotide encoding the protein can also be inserted into a suitable insect expression vector, and is operably linked to the control elements within that vector. Vector construction employs techniques which are known in the art. Generally, the components of the expression system include a transfer vector, usually a bacterial plasmid, which contains both a fragment of the baculovirus

genome, and a convenient restriction site for insertion of the heterologous gene or genes to be expressed; a wild type baculovirus with a sequence homologous to the baculovirus-specific fragment in the transfer vector (this allows for the homologous recombination of the heterologous gene in to the baculovirus genome); and appropriate insect host cells and growth media.

- 5 After inserting the DNA sequence encoding the protein into the transfer vector, the vector and the wild type viral genome are transfected into an insect host cell where the vector and viral genome are allowed to recombine. The packaged recombinant virus is expressed and recombinant plaques are identified and purified. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, *inter alia*, Invitrogen, San Diego CA ("MaxBac" kit).
- 10 These techniques are generally known to those skilled in the art and fully described in Summers and Smith, *Texas Agricultural Experiment Station Bulletin No. 1555* (1987) (hereinafter "Summers and Smith").

- Prior to inserting the DNA sequence encoding the protein into the baculovirus genome, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and
- 15 transcription termination sequence, are usually assembled into an intermediate transplacement construct (transfer vector). This construct may contain a single gene and operably linked regulatory elements; multiple genes, each with its own set of operably linked regulatory elements; or multiple genes, regulated by the same set of regulatory elements. Intermediate transplacement constructs are often maintained in a replicon, such as an extrachromosomal element (*eg.* plasmids) capable of stable
 - 20 maintenance in a host, such as a bacterium. The replicon will have a replication system, thus allowing it to be maintained in a suitable host for cloning and amplification.

- Currently, the most commonly used transfer vector for introducing foreign genes into AcNPV is pAc373. Many other vectors, known to those of skill in the art, have also been designed. These include, for example, pVL985 (which alters the polyhedrin start codon from ATG to ATT, and
- 25 which introduces a BamHI cloning site 32 basepairs downstream from the ATT; see Luckow and Summers, *Virology* (1989) 17:31.

The plasmid usually also contains the polyhedrin polyadenylation signal (Miller et al. (1988) *Ann. Rev. Microbiol.*, 42:177) and a prokaryotic ampicillin-resistance (*amp*) gene and origin of replication for selection and propagation in *E. coli*.

Baculovirus transfer vectors usually contain a baculovirus promoter. A baculovirus promoter is any DNA sequence capable of binding a baculovirus RNA polymerase and initiating the downstream (5' to 3') transcription of a coding sequence (eg. structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A baculovirus transfer vector may also have a second domain called an enhancer, which, if present, is usually distal to the structural gene. Expression may be either regulated or constitutive.

Structural genes, abundantly transcribed at late times in a viral infection cycle, provide particularly useful promoter sequences. Examples include sequences derived from the gene encoding the viral polyhedron protein, Friesen et al., (1986) "The Regulation of Baculovirus Gene Expression," in: *The Molecular Biology of Baculoviruses* (ed. Walter Doerfler); EPO Publ. Nos. 127 839 and 155 476; and the gene encoding the p10 protein, Vlcek et al., (1988), *J. Gen. Virol.* 69:765.

DNA encoding suitable signal sequences can be derived from genes for secreted insect or baculovirus proteins, such as the baculovirus polyhedrin gene (Carbonell et al. (1988) *Gene*, 73:409). Alternatively, since the signals for mammalian cell posttranslational modifications (such as signal peptide cleavage, proteolytic cleavage, and phosphorylation) appear to be recognized by insect cells, and the signals required for secretion and nuclear accumulation also appear to be conserved between the invertebrate cells and vertebrate cells, leaders of non-insect origin, such as those derived from genes encoding human α -interferon, Maeda et al., (1985), *Nature* 315:592; human gastrin-releasing peptide, Lebacqz-Verheyden et al., (1988), *Molec. Cell. Biol.* 8:3129; human IL-2, Smith et al., (1985) *Proc. Nat'l Acad. Sci. USA*, 82:8404; mouse IL-3, (Miyajima et al., (1987) *Gene* 58:273; and human glucocerebrosidase, Martin et al. (1988) *DNA*, 7:99, can also be used to provide for secretion in insects.

A recombinant polypeptide or polyprotein may be expressed intracellularly or, if it is expressed with the proper regulatory sequences, it can be secreted. Good intracellular expression of nonfused foreign proteins usually requires heterologous genes that ideally have a short leader sequence containing suitable translation initiation signals preceding an ATG start signal. If desired, methionine at the N-terminus may be cleaved from the mature protein by *in vitro* incubation with cyanogen bromide.

Alternatively, recombinant polypeptides or proteins which are not naturally secreted can be secreted from the insect cell by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provides for secretion of the foreign protein in insects. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the translocation of the protein into the endoplasmic reticulum.

After insertion of the DNA sequence and/or the gene encoding the expression product precursor of the protein, an insect cell host is co-transformed with the heterologous DNA of the transfer vector and the genomic DNA of wild type baculovirus -- usually by co-transfection. The promoter and transcription termination sequence of the construct will usually comprise a 2-5kb section of the baculovirus genome. Methods for introducing heterologous DNA into the desired site in the baculovirus virus are known in the art. (See Summers and Smith *supra*; Ju et al. (1987); Smith et al., *Mol. Cell. Biol.* (1983) 3:2156; and Luckow and Summers (1989)). For example, the insertion can be into a gene such as the polyhedrin gene, by homologous double crossover recombination; insertion can also be into a restriction enzyme site engineered into the desired baculovirus gene. Miller et al., (1989), *Bioessays* 4:91. The DNA sequence, when cloned in place of the polyhedrin gene in the expression vector, is flanked both 5' and 3' by polyhedrin-specific sequences and is positioned downstream of the polyhedrin promoter.

The newly formed baculovirus expression vector is subsequently packaged into an infectious recombinant baculovirus. Homologous recombination occurs at low frequency (between about 1% and about 5%); thus, the majority of the virus produced after cotransfection is still wild-type virus. Therefore, a method is necessary to identify recombinant viruses. An advantage of the expression system is a visual screen allowing recombinant viruses to be distinguished. The polyhedrin protein, which is produced by the native virus, is produced at very high levels in the nuclei of infected cells at late times after viral infection. Accumulated polyhedrin protein forms occlusion bodies that also contain embedded particles. These occlusion bodies, up to 15 μ m in size, are highly refractile, giving them a bright shiny appearance that is readily visualized under the light microscope. Cells infected with recombinant viruses lack occlusion bodies. To distinguish recombinant virus from wild-type virus, the transfection supernatant is plaqued onto a monolayer of insect cells by techniques known to those skilled in the art. Namely, the plaques are screened under the light microscope for the presence (indicative of wild-type virus) or absence (indicative of recombinant

virus) of occlusion bodies. "Current Protocols in Microbiology" Vol. 2 (Ausubel et al. eds) at 16.8 (Supp. 10, 1990); Summers and Smith, *supra*; Miller et al. (1989).

Recombinant baculovirus expression vectors have been developed for infection into several insect cells. For example, recombinant baculoviruses have been developed for, *inter alia*: *Aedes aegypti*, *Autographa californica*, *Bombyx mori*, *Drosophila melanogaster*, *Spodoptera frugiperda*, and *Trichoplusia ni* (WO 89/046699; Carbonell et al., (1985) *J. Virol.* 56:153; Wright (1986) *Nature* 321:718; Smith et al., (1983) *Mol. Cell. Biol.* 3:2156; and see generally, Fraser, *et al.* (1989) *In Vitro Cell. Dev. Biol.* 25:225).

Cells and cell culture media are commercially available for both direct and fusion expression of heterologous polypeptides in a baculovirus/expression system; cell culture technology is generally known to those skilled in the art. *See, eg.* Summers and Smith *supra*.

The modified insect cells may then be grown in an appropriate nutrient medium, which allows for stable maintenance of the plasmid(s) present in the modified insect host. Where the expression product gene is under inducible control, the host may be grown to high density, and expression induced. Alternatively, where expression is constitutive, the product will be continuously expressed into the medium and the nutrient medium must be continuously circulated, while removing the product of interest and augmenting depleted nutrients. The product may be purified by such techniques as chromatography, *eg.* HPLC, affinity chromatography, ion exchange chromatography, *etc.*; electrophoresis; density gradient centrifugation; solvent extraction, or the like. As appropriate, the product may be further purified, as required, so as to remove substantially any insect proteins which are also secreted in the medium or result from lysis of insect cells, so as to provide a product which is at least substantially free of host debris, *eg.* proteins, lipids and polysaccharides.

In order to obtain protein expression, recombinant host cells derived from the transformants are incubated under conditions which allow expression of the recombinant protein encoding sequence.

These conditions will vary, dependent upon the host cell selected. However, the conditions are readily ascertainable to those of ordinary skill in the art, based upon what is known in the art.

iii. Plant Systems

There are many plant cell culture and whole plant genetic expression systems known in the art. Exemplary plant cellular genetic expression systems include those described in patents, such as:

US 5,693,506; US 5,659,122; and US 5,608,143. Additional examples of genetic expression in plant cell culture has been described by Zenk, *Phytochemistry* 30:3861-3863 (1991). Descriptions of plant protein signal peptides may be found in addition to the references described above in Vaulcombe et al., *Mol. Gen. Genet.* 209:33-40 (1987); Chandler et al., *Plant Molecular Biology* 3:407-418 (1984); Rogers, *J. Biol. Chem.* 260:3731-3738 (1985); Rothstein et al., *Gene* 55:353-356 (1987); Whittier et al., *Nucleic Acids Research* 15:2515-2535 (1987); Wirsal et al., *Molecular Microbiology* 3:3-14 (1989); Yu et al., *Gene* 122:247-253 (1992). A description of the regulation of plant gene expression by the phytohormone, gibberellic acid and secreted enzymes induced by gibberellic acid can be found in R.L. Jones and J. MacMillin, *Gibberellins*: in: *Advanced Plant Physiology*, Malcolm B. Wilkins, ed., 1984 Pitman Publishing Limited, London, pp. 21-52. References that describe other metabolically-regulated genes: Sheen, *Plant Cell*, 2:1027-1038(1990); Maas et al., *EMBO J.* 9:3447-3452 (1990); Benkel and Hickey, *Proc. Natl. Acad. Sci.* 84:1337-1339 (1987)

Typically, using techniques known in the art, a desired polynucleotide sequence is inserted into an expression cassette comprising genetic regulatory elements designed for operation in plants. The expression cassette is inserted into a desired expression vector with companion sequences upstream and downstream from the expression cassette suitable for expression in a plant host. The companion sequences will be of plasmid or viral origin and provide necessary characteristics to the vector to permit the vectors to move DNA from an original cloning host, such as bacteria, to the desired plant host. The basic bacterial/plant vector construct will preferably provide a broad host range prokaryote replication origin; a prokaryote selectable marker; and, for *Agrobacterium* transformations, T DNA sequences for *Agrobacterium*-mediated transfer to plant chromosomes. Where the heterologous gene is not readily amenable to detection, the construct will preferably also have a selectable marker gene suitable for determining if a plant cell has been transformed. A general review of suitable markers, for example for the members of the grass family, is found in Wilmink and Dons, 1993, *Plant Mol. Biol. Repr*, 11(2):165-185.

Sequences suitable for permitting integration of the heterologous sequence into the plant genome are also recommended. These might include transposon sequences and the like for homologous recombination as well as Ti sequences which permit random insertion of a heterologous expression cassette into a plant genome. Suitable prokaryote selectable markers include resistance toward

antibiotics such as ampicillin or tetracycline. Other DNA sequences encoding additional functions may also be present in the vector, as is known in the art.

The nucleic acid molecules of the subject invention may be included into an expression cassette for expression of the protein(s) of interest. Usually, there will be only one expression cassette, although two or more are feasible. The recombinant expression cassette will contain in addition to the heterologous protein encoding sequence the following elements, a promoter region, plant 5' untranslated sequences, initiation codon depending upon whether or not the structural gene comes equipped with one, and a transcription and translation termination sequence. Unique restriction enzyme sites at the 5' and 3' ends of the cassette allow for easy insertion into a pre-existing vector.

10 A heterologous coding sequence may be for any protein relating to the present invention. The sequence encoding the protein of interest will encode a signal peptide which allows processing and translocation of the protein, as appropriate, and will usually lack any sequence which might result in the binding of the desired protein of the invention to a membrane. Since, for the most part, the transcriptional initiation region will be for a gene which is expressed and translocated during
15 germination, by employing the signal peptide which provides for translocation, one may also provide for translocation of the protein of interest. In this way, the protein(s) of interest will be translocated from the cells in which they are expressed and may be efficiently harvested. Typically secretion in seeds are across the aleurone or scutellar epithelium layer into the endosperm of the seed. While it is not required that the protein be secreted from the cells in which the protein is
20 produced, this facilitates the isolation and purification of the recombinant protein.

Since the ultimate expression of the desired gene product will be in a eucaryotic cell it is desirable to determine whether any portion of the cloned gene contains sequences which will be processed out as introns by the host's splicosome machinery. If so, site-directed mutagenesis of the "intron" region may be conducted to prevent losing a portion of the genetic message as a false intron code,
25 Reed and Maniatis, *Cell* 41:95-105, 1985.

The vector can be microinjected directly into plant cells by use of micropipettes to mechanically transfer the recombinant DNA. Crossway, *Mol. Gen. Genet*, 202:179-185, 1985. The genetic material may also be transferred into the plant cell by using polyethylene glycol, Krens, et al., *Nature*, 296, 72-74, 1982. Another method of introduction of nucleic acid segments is high

velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads or particles, or on the surface, Klein, et al., *Nature*, 327, 70-73, 1987 and Knudsen and Muller, 1991, *Planta*, 185:330-336 teaching particle bombardment of barley endosperm to create transgenic barley. Yet another method of introduction would be fusion of protoplasts with other entities, either minicells, cells, lysosomes or other fusible lipid-surfaced bodies, Fraley, et al., *Proc. Natl. Acad. Sci. USA*, 79, 1859-1863, 1982.

The vector may also be introduced into the plant cells by electroporation. (Fromm et al., *Proc. Natl. Acad. Sci. USA* 82:5824, 1985). In this technique, plant protoplasts are electroporated in the presence of plasmids containing the gene construct. Electrical impulses of high field strength reversibly permeabilize biomembranes allowing the introduction of the plasmids. Electroporated plant protoplasts reform the cell wall, divide, and form plant callus.

All plants from which protoplasts can be isolated and cultured to give whole regenerated plants can be transformed by the present invention so that whole plants are recovered which contain the transferred gene. It is known that practically all plants can be regenerated from cultured cells or tissues, including but not limited to all major species of sugarcane, sugar beet, cotton, fruit and other trees, legumes and vegetables. Some suitable plants include, for example, species from the genera *Fragaria*, *Lotus*, *Medicago*, *Onobrychis*, *Trifolium*, *Trigonella*, *Vigna*, *Citrus*, *Linum*, *Geranium*, *Manihot*, *Daucus*, *Arabidopsis*, *Brassica*, *Raphanus*, *Sinapis*, *Atropa*, *Capsicum*, *Datura*, *Hyoscyamus*, *Lycopersion*, *Nicotiana*, *Solanum*, *Petunia*, *Digitalis*, *Majorana*, *Cichorium*, *Helianthus*, *Lactuca*, *Bromus*, *Asparagus*, *Antirrhinum*, *Hererocallis*, *Nemesia*, *Pelargonium*, *Panicum*, *Pennisetum*, *Ranunculus*, *Senecio*, *Salpiglossis*, *Cucumis*, *Browaalia*, *Glycine*, *Lolium*, *Zea*, *Triticum*, *Sorghum*, and *Datura*.

Means for regeneration vary from species to species of plants, but generally a suspension of transformed protoplasts containing copies of the heterologous gene is first provided. Callus tissue is formed and shoots may be induced from callus and subsequently rooted. Alternatively, embryo formation can be induced from the protoplast suspension. These embryos germinate as natural embryos to form plants. The culture media will generally contain various amino acids and hormones, such as auxin and cytokinins. It is also advantageous to add glutamic acid and proline to the medium, especially for such species as corn and alfalfa. Shoots and roots normally develop simultaneously. Efficient regeneration will depend on the medium, on the genotype, and on the

history of the culture. If these three variables are controlled, then regeneration is fully reproducible and repeatable.

In some plant cell culture systems, the desired protein of the invention may be excreted or alternatively, the protein may be extracted from the whole plant. Where the desired protein of the invention is secreted into the medium, it may be collected. Alternatively, the embryos and embryoless-half seeds or other plant tissue may be mechanically disrupted to release any secreted protein between cells and tissues. The mixture may be suspended in a buffer solution to retrieve soluble proteins. Conventional protein isolation and purification methods will be then used to purify the recombinant protein. Parameters of time, temperature pH, oxygen, and volumes will be adjusted through routine methods to optimize expression and recovery of heterologous protein.

iv. Bacterial Systems

Bacterial expression techniques are known in the art. A bacterial promoter is any DNA sequence capable of binding bacterial RNA polymerase and initiating the downstream (3') transcription of a coding sequence (eg. structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A bacterial promoter may also have a second domain called an operator, that may overlap an adjacent RNA polymerase binding site at which RNA synthesis begins. The operator permits negative regulated (inducible) transcription, as a gene repressor protein may bind the operator and thereby inhibit transcription of a specific gene. Constitutive expression may occur in the absence of negative regulatory elements, such as the operator. In addition, positive regulation may be achieved by a gene activator protein binding sequence, which, if present is usually proximal (5') to the RNA polymerase binding sequence. An example of a gene activator protein is the catabolite activator protein (CAP), which helps initiate transcription of the lac operon in *Escherichia coli* (*E. coli*) [Raibaud *et al.* (1984) *Annu. Rev. Genet.* 18:173]. Regulated expression may therefore be either positive or negative, thereby either enhancing or reducing transcription.

Sequences encoding metabolic pathway enzymes provide particularly useful promoter sequences. Examples include promoter sequences derived from sugar metabolizing enzymes, such as galactose, lactose (*lac*) [Chang *et al.* (1977) *Nature* 198:1056], and maltose. Additional examples include promoter sequences derived from biosynthetic enzymes such as tryptophan (*trp*) [Goeddel *et al.*

(1980) *Nuc. Acids Res.* 8:4057; Yelverton *et al.* (1981) *Nucl. Acids Res.* 9:731; US patent 4,738,921; EP-A-0036776 and EP-A-0121775]. The g-laotamase (*bla*) promoter system [Weissmann (1981) "The cloning of interferon and other mistakes." In *Interferon 3* (ed. I. Gresser)], bacteriophage lambda PL [Shimatake *et al.* (1981) *Nature* 292:128] and T5 [US patent 4,689,406]
5 promoter systems also provide useful promoter sequences.

In addition, synthetic promoters which do not occur in nature also function as bacterial promoters. For example, transcription activation sequences of one bacterial or bacteriophage promoter may be joined with the operon sequences of another bacterial or bacteriophage promoter, creating a synthetic hybrid promoter [US patent 4,551,433]. For example, the *tac* promoter is a hybrid *trp-lac*
10 promoter comprised of both *trp* promoter and *lac* operon sequences that is regulated by the *lac* repressor [Amann *et al.* (1983) *Gene* 25:167; de Boer *et al.* (1983) *Proc. Natl. Acad. Sci.* 80:21]. Furthermore, a bacterial promoter can include naturally occurring promoters of non-bacterial origin that have the ability to bind bacterial RNA polymerase and initiate transcription. A naturally occurring promoter of non-bacterial origin can also be coupled with a compatible RNA polymerase
15 to produce high levels of expression of some genes in prokaryotes. The bacteriophage T7 RNA polymerase/promoter system is an example of a coupled promoter system [Studier *et al.* (1986) *J. Mol. Biol.* 189:113; Tabor *et al.* (1985) *Proc Natl. Acad. Sci.* 82:1074]. In addition, a hybrid promoter can also be comprised of a bacteriophage promoter and an *E. coli* operator region (EPO-A-0 267 851).

20 In addition to a functioning promoter sequence, an efficient ribosome binding site is also useful for the expression of foreign genes in prokaryotes. In *E. coli*, the ribosome binding site is called the Shine-Dalgarno (SD) sequence and includes an initiation codon (ATG) and a sequence 3-9 nucleotides in length located 3-11 nucleotides upstream of the initiation codon [Shine *et al.* (1975) *Nature* 254:34]. The SD sequence is thought to promote binding of mRNA to the ribosome by the
25 pairing of bases between the SD sequence and the 3' end of *E. coli* 16S rRNA [Steitz *et al.* (1979) "Genetic signals and nucleotide sequences in messenger RNA." In *Biological Regulation and Development: Gene Expression* (ed. R.F. Goldberger)]. To express eukaryotic genes and prokaryotic genes with weak ribosome-binding site [Sambrook *et al.* (1989) "Expression of cloned genes in *Escherichia coli*." In *Molecular Cloning: A Laboratory Manual*].

A DNA molecule may be expressed intracellularly. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus will always be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide or by either *in vivo* on *in vitro* incubation with a bacterial methionine N-terminal peptidase (EPO-A-0 219 237).

Fusion proteins provide an alternative to direct expression. Usually, a DNA sequence encoding the N-terminal portion of an endogenous bacterial protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the bacteriophage lambda cell gene can be linked at the 5' terminus of a foreign gene and expressed in bacteria. The resulting fusion protein preferably retains a site for a processing enzyme (factor Xa) to cleave the bacteriophage protein from the foreign gene [Nagai *et al.* (1984) *Nature* 309:810]. Fusion proteins can also be made with sequences from the *lacZ* [Jia *et al.* (1987) *Gene* 60:197], *trpE* [Allen *et al.* (1987) *J. Biotechnol.* 5:93; Makoff *et al.* (1989) *J. Gen. Microbiol.* 135:11], and *Chey* [EP-A-0 324 647] genes. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (*eg.* ubiquitin specific processing-protease) to cleave the ubiquitin from the foreign protein. Through this method, native foreign protein can be isolated [Miller *et al.* (1989) *Bio/Technology* 7:698].

Alternatively, foreign proteins can also be secreted from the cell by creating chimeric DNA molecules that encode a fusion protein comprised of a signal peptide sequence fragment that provides for secretion of the foreign protein in bacteria [US patent 4,336,336]. The signal sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The protein is either secreted into the growth media (gram-positive bacteria) or into the periplasmic space, located between the inner and outer membrane of the cell (gram-negative bacteria). Preferably there are processing sites, which can be cleaved either *in vivo* or *in vitro* encoded between the signal peptide fragment and the foreign gene.

DNA encoding suitable signal sequences can be derived from genes for secreted bacterial proteins, such as the *E. coli* outer membrane protein gene (*ompA*) [Masui *et al.* (1983), in: *Experimental Manipulation of Gene Expression*; Ghrayeb *et al.* (1984) *EMBO J.* 3:2437] and the *E. coli* alkaline

phosphatase signal sequence (*phoA*) [Oka *et al.* (1985) *Proc. Natl. Acad. Sci.* 82:7212]. As an additional example, the signal sequence of the alpha-amylase gene from various *Bacillus* strains can be used to secrete heterologous proteins from *B. subtilis* [Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 244 042].

- 5 Usually, transcription termination sequences recognized by bacteria are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Transcription termination sequences frequently include DNA sequences of about 50 nucleotides capable of forming stem loop structures that aid in terminating transcription.
- 10 Examples include transcription termination sequences derived from genes with strong promoters, such as the *trp* gene in *E. coli* as well as other biosynthetic genes.

Usually, the above described components, comprising a promoter, signal sequence (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal

15 element (eg. plasmids) capable of stable maintenance in a host, such as bacteria. The replicon will have a replication system, thus allowing it to be maintained in a prokaryotic host either for expression or for cloning and amplification. In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy

20 number plasmid will preferably contain at least about 10, and more preferably at least about 20 plasmids. Either a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host.

Alternatively, the expression constructs can be integrated into the bacterial genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to the

25 bacterial chromosome that allows the vector to integrate. Integrations appear to result from recombinations between homologous DNA in the vector and the bacterial chromosome. For example, integrating vectors constructed with DNA from various *Bacillus* strains integrate into the *Bacillus* chromosome (EP-A- 0 127 328). Integrating vectors may also be comprised of bacteriophage or transposon sequences.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of bacterial strains that have been transformed. Selectable markers can be expressed in the bacterial host and may include genes which render bacteria resistant to drugs such as ampicillin, chloramphenicol, erythromycin, kanamycin (neomycin), and tetracycline [Davies *et al.* (1978) *Annu. Rev. Microbiol.* 32:469]. Selectable markers may also include biosynthetic genes, such as those in the histidine, tryptophan, and leucine biosynthetic pathways.

Alternatively, some of the above described components can be put together in transformation vectors. Transformation vectors are usually comprised of a selectable marker that is either maintained in a replicon or developed into an integrating vector, as described above.

Expression and transformation vectors, either extra-chromosomal replicons or integrating vectors, have been developed for transformation into many bacteria. For example, expression vectors have been developed for, *inter alia*, the following bacteria: *Bacillus subtilis* [Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541], *Escherichia coli* [Shimatake *et al.* (1981) *Nature* 292:128; Amann *et al.* (1985) *Gene* 40:183; Studier *et al.* (1986) *J. Mol. Biol.* 189:113; EP-A-0 036 776, EP-A-0 136 829 and EP-A-0 136 907], *Streptococcus cremoris* [Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655], *Streptococcus lividans* [Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655], *Streptomyces lividans* [US patent 4,745,056].

Methods of introducing exogenous DNA into bacterial hosts are well-known in the art, and usually include either the transformation of bacteria treated with CaCl_2 or other agents, such as divalent cations and DMSO. DNA can also be introduced into bacterial cells by electroporation. Transformation procedures usually vary with the bacterial species to be transformed. See eg. [Masson *et al.* (1989) *FEMS Microbiol. Lett.* 60:273; Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541, *Bacillus*], [Miller *et al.* (1988) *Proc. Natl. Acad. Sci.* 85:856; Wang *et al.* (1990) *J. Bacteriol.* 172:949, *Campylobacter*], [Cohen *et al.* (1973) *Proc. Natl. Acad. Sci.* 69:2110; Dower *et al.* (1988) *Nucleic Acids Res.* 16:6127; Kushner (1978) "An improved method for transformation of *Escherichia coli* with ColE1-derived plasmids. In *Genetic Engineering: Proceedings of the International Symposium on Genetic Engineering* (eds. H.W. Boyer and S. Nicosia); Mandel *et al.* (1970) *J. Mol. Biol.* 53:159; Taketo (1988) *Biochim. Biophys. Acta* 949:318; *Escherichia*], [Chassy *et al.* (1987) *FEMS Microbiol. Lett.*

- 44:173 *Lactobacillus*]; [Fiedler *et al.* (1988) *Anal. Biochem* 170:38, *Pseudomonas*]; [Augustin *et al.* (1990) *FEMS Microbiol. Lett.* 66:203, *Staphylococcus*], [Barany *et al.* (1980) *J. Bacteriol.* 144:698; Harlander (1987) "Transformation of *Streptococcus lactis* by electroporation, in: *Streptococcal Genetics* (ed. J. Ferretti and R. Curtiss III); Perry *et al.* (1981) *Infect. Immun.* 32:1295; Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655; Somkuti *et al.* (1987) *Proc. 4th Eur. Cong. Biotechnology* 1:412, *Streptococcus*].

v. Yeast Expression

- Yeast expression systems are also known to one of ordinary skill in the art. A yeast promoter is any DNA sequence capable of binding yeast RNA polymerase and initiating the downstream (3') transcription of a coding sequence (*eg.* structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site (the "TATA Box") and a transcription initiation site. A yeast promoter may also have a second domain called an upstream activator sequence (UAS), which, if present, is usually distal to the structural gene. The UAS permits regulated (inducible) expression. Constitutive expression occurs in the absence of a UAS. Regulated expression may be either positive or negative, thereby either enhancing or reducing transcription.

- Yeast is a fermenting organism with an active metabolic pathway, therefore sequences encoding enzymes in the metabolic pathway provide particularly useful promoter sequences. Examples include alcohol dehydrogenase (ADH) (EP-A-0 284 044), enolase, glucokinase, glucose-6-phosphate isomerase, glyceraldehyde-3-phosphate-dehydrogenase (GAP or GAPDH), hexokinase, phosphofructokinase, 3-phosphoglycerate mutase, and pyruvate kinase (PyK) (EPO-A-0 329 203). The yeast *PHO5* gene, encoding acid phosphatase, also provides useful promoter sequences [Myanohara *et al.* (1983) *Proc. Natl. Acad. Sci. USA* 80:1].

- In addition, synthetic promoters which do not occur in nature also function as yeast promoters. For example, UAS sequences of one yeast promoter may be joined with the transcription activation region of another yeast promoter, creating a synthetic hybrid promoter. Examples of such hybrid promoters include the ADH regulatory sequence linked to the GAP transcription activation region (US Patent Nos. 4,876,197 and 4,880,734). Other examples of hybrid promoters include promoters which consist of the regulatory sequences of either the *ADH2*, *GAL4*, *GAL10*, OR *PHO5* genes,

combined with the transcriptional activation region of a glycolytic enzyme gene such as GAP or PyK (EP-A-0 164 556). Furthermore, a yeast promoter can include naturally occurring promoters of non-yeast origin that have the ability to bind yeast RNA polymerase and initiate transcription. Examples of such promoters include, *inter alia*, [Cohen *et al.* (1980) *Proc. Natl. Acad. Sci. USA* 77:1078; Henikoff *et al.* (1981) *Nature* 283:835; Hollenberg *et al.* (1981) *Curr. Topics Microbiol. Immunol.* 96:119; Hollenberg *et al.* (1979) "The Expression of Bacterial Antibiotic Resistance Genes in the Yeast *Saccharomyces cerevisiae*," in: *Plasmids of Medical, Environmental and Commercial Importance* (eds. K.N. Timmis and A. Puhler); Mercerau-Puigalon *et al.* (1980) *Gene* 11:163; Panthier *et al.* (1980) *Curr. Genet.* 2:109;].

- 10 A DNA molecule may be expressed intracellularly in yeast. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide.
- 15 Fusion proteins provide an alternative for yeast expression systems, as well as in mammalian, baculovirus, and bacterial expression systems. Usually, a DNA sequence encoding the N-terminal portion of an endogenous yeast protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the yeast or human superoxide dismutase (SOD) gene, can be
- 20 linked at the 5' terminus of a foreign gene and expressed in yeast. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. See *eg.* EP-A-0 196 056. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (*eg.* ubiquitin-specific processing protease) to cleave the ubiquitin from the foreign protein. Through this method,
- 25 therefore, native foreign protein can be isolated (*eg.* WO88/024066).

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provide for secretion in yeast of the foreign protein. Preferably, there are processing sites encoded between the leader fragment and the foreign gene that can be cleaved either *in vivo* or *in vitro*. The

leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell.

DNA encoding suitable signal sequences can be derived from genes for secreted yeast proteins, such as the yeast invertase gene (EP-A-0 012 873; JPO. 62,096,086) and the A-factor gene (US
5 patent 4,588,684). Alternatively, leaders of non-yeast origin, such as an interferon leader, exist that also provide for secretion in yeast (EP-A-0 060 057).

A preferred class of secretion leaders are those that employ a fragment of the yeast alpha-factor gene, which contains both a "pre" signal sequence, and a "pro" region. The types of alpha-factor fragments that can be employed include the full-length pre-pro alpha factor leader (about 83 amino
10 acid residues) as well as truncated alpha-factor leaders (usually about 25 to about 50 amino acid residues) (US Patents 4,546,083 and 4,870,008; EP-A-0 324 274). Additional leaders employing an alpha-factor leader fragment that provides for secretion include hybrid alpha-factor leaders made with a presequence of a first yeast, but a pro-region from a second yeast alphafactor. (eg. see WO 89/02463.)

15 Usually, transcription termination sequences recognized by yeast are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator sequence and other yeast-recognized termination sequences, such as those coding for glycolytic enzymes.

20 Usually, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (eg. plasmids) capable of stable maintenance in a host, such as yeast or bacteria. The replicon may have two replication systems, thus allowing it to be maintained, for example, in yeast
25 for expression and in a prokaryotic host for cloning and amplification. Examples of such yeast-bacteria shuttle vectors include YEp24 [Botstein *et al.* (1979) *Gene* 8:17-24], pCI/1 [Brake *et al.* (1984) *Proc. Natl. Acad. Sci USA* 81:4642-4646], and YRp17 [Stinchcomb *et al.* (1982) *J. Mol. Biol.* 158:157]. In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number ranging from about 5 to about 200, and

usually about 10 to about 150. A host containing a high copy number plasmid will preferably have at least about 10, and more preferably at least about 20. Enter a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host. See *eg. Brake et al., supra*.

- 5 Alternatively, the expression constructs can be integrated into the yeast genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to a yeast chromosome that allows the vector to integrate, and preferably contain two homologous sequences flanking the expression construct. Integrations appear to result from recombinations between homologous DNA in the vector and the yeast chromosome [Orr-Weaver *et al.* (1983) *Methods in*
10 *Enzymol.* 101:228-245]. An integrating vector may be directed to a specific locus in yeast by selecting the appropriate homologous sequence for inclusion in the vector. See Orr-Weaver *et al., supra*. One or more expression construct may integrate, possibly affecting levels of recombinant protein produced [Rine *et al.* (1983) *Proc. Natl. Acad. Sci. USA* 80:6750]. The chromosomal sequences included in the vector can occur either as a single segment in the vector, which results
15 in the integration of the entire vector, or two segments homologous to adjacent segments in the chromosome and flanking the expression construct in the vector, which can result in the stable integration of only the expression construct.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of yeast strains that have been transformed. Selectable markers may
20 include biosynthetic genes that can be expressed in the yeast host, such as *ADE2*, *HIS4*, *LEU2*, *TRP1*, and *ALG7*, and the G418 resistance gene, which confer resistance in yeast cells to tunicamycin and G418, respectively. In addition, a suitable selectable marker may also provide yeast with the ability to grow in the presence of toxic compounds, such as metal. For example, the presence of *CUP1* allows yeast to grow in the presence of copper ions [Butt *et al.* (1987) *Microbiol.*
25 *Rev.* 51:351].

Alternatively, some of the above described components can be put together into transformation vectors. Transformation vectors are usually comprised of a selectable marker that is either maintained in a replicon or developed into an integrating vector, as described above.

Expression and transformation vectors, either extrachromosomal replicons or integrating vectors, have been developed for transformation into many yeasts. For example, expression vectors have been developed for, *inter alia*, the following yeasts: *Candida albicans* [Kurtz, *et al.* (1986) *Mol. Cell. Biol.* 6:142], *Candida maltosa* [Kunze, *et al.* (1985) *J. Basic Microbiol.* 25:141], *Hansenula polymorpha* [Gleeson, *et al.* (1986) *J. Gen. Microbiol.* 132:3459; Roggenkamp *et al.* (1986) *Mol. Gen. Genet.* 202:302], *Kluyveromyces fragilis* [Das, *et al.* (1984) *J. Bacteriol.* 158:1165], *Kluyveromyces lactis* [De Louvencourt *et al.* (1983) *J. Bacteriol.* 154:737; Van den Berg *et al.* (1990) *Bio/Technology* 8:135], *Pichia guillerimondii* [Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141], *Pichia pastoris* [Cregg, *et al.* (1985) *Mol. Cell. Biol.* 5:3376; US Patent Nos. 4,837,148 and 4,929,555], *Saccharomyces cerevisiae* [Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci. USA* 75:1929; Ito *et al.* (1983) *J. Bacteriol.* 153:163], *Schizosaccharomyces pombe* [Beach and Nurse (1981) *Nature* 300:706], and *Yarrowia lipolytica* [Davidow, *et al.* (1985) *Curr. Genet.* 10:380471 Gaillardin, *et al.* (1985) *Curr. Genet.* 10:49].

Methods of introducing exogenous DNA into yeast hosts are well-known in the art, and usually include either the transformation of spheroplasts or of intact yeast cells treated with alkali cations. Transformation procedures usually vary with the yeast species to be transformed. See *eg.* [Kurtz *et al.* (1986) *Mol. Cell. Biol.* 6:142; Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141; *Candida*]; [Gleeson *et al.* (1986) *J. Gen. Microbiol.* 132:3459; Roggenkamp *et al.* (1986) *Mol. Gen. Genet.* 202:302; *Hansenula*]; [Das *et al.* (1984) *J. Bacteriol.* 158:1165; De Louvencourt *et al.* (1983) *J. Bacteriol.* 154:1165; Van den Berg *et al.* (1990) *Bio/Technology* 8:135; *Kluyveromyces*]; [Cregg *et al.* (1985) *Mol. Cell. Biol.* 5:3376; Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141; US Patent Nos. 4,837,148 and 4,929,555; *Pichia*]; [Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci. USA* 75:1929; Ito *et al.* (1983) *J. Bacteriol.* 153:163 *Saccharomyces*]; [Beach and Nurse (1981) *Nature* 300:706; *Schizosaccharomyces*]; [Davidow *et al.* (1985) *Curr. Genet.* 10:39; Gaillardin *et al.* (1985) *Curr. Genet.* 10:49; *Yarrowia*].

Antibodies

As used herein, the term "antibody" refers to a polypeptide or group of polypeptides composed of at least one antibody combining site. An "antibody combining site" is the three-dimensional binding space with an internal surface shape and charge distribution complementary to the features of an epitope of an antigen, which allows a binding of the antibody with the antigen. "Antibody"

includes, for example, vertebrate antibodies, hybrid antibodies, chimeric antibodies, humanised antibodies, altered antibodies, univalent antibodies, Fab proteins, and single domain antibodies.

Antibodies against the proteins of the invention are useful for affinity chromatography, immunoassays, and distinguishing/identifying Neisserial proteins.

- 5 Antibodies to the proteins of the invention, both polyclonal and monoclonal, may be prepared by conventional methods. In general, the protein is first used to immunize a suitable animal, preferably a mouse, rat, rabbit or goat. Rabbits and goats are preferred for the preparation of polyclonal sera due to the volume of serum obtainable, and the availability of labeled anti-rabbit and anti-goat antibodies. Immunization is generally performed by mixing or emulsifying the protein in saline, preferably in an adjuvant such as Freund's complete adjuvant, and injecting the mixture or emulsion parenterally (generally subcutaneously or intramuscularly). A dose of 50-200 µg/injection is typically sufficient. Immunization is generally boosted 2-6 weeks later with one or more injections of the protein in saline, preferably using Freund's incomplete adjuvant. One may alternatively generate antibodies by *in vitro* immunization using methods known in the art, which for the purposes of this invention is considered equivalent to *in vivo* immunization. Polyclonal antisera is obtained by bleeding the immunized animal into a glass or plastic container, incubating the blood at 25°C for one hour, followed by incubating at 4°C for 2-18 hours. The serum is recovered by centrifugation (eg. 1,000g for 10 minutes). About 20-50 ml per bleed may be obtained from rabbits.
- 10 20 Monoclonal antibodies are prepared using the standard method of Kohler & Milstein [*Nature* (1975) 256:495-96], or a modification thereof. Typically, a mouse or rat is immunized as described above. However, rather than bleeding the animal to extract serum, the spleen (and optionally several large lymph nodes) is removed and dissociated into single cells. If desired, the spleen cells may be screened (after removal of nonspecifically adherent cells) by applying a cell suspension to a plate or well coated with the protein antigen. B-cells expressing membrane-bound immunoglobulin specific for the antigen bind to the plate, and are not rinsed away with the rest of the suspension. Resulting B-cells, or all dissociated spleen cells, are then induced to fuse with myeloma cells to form hybridomas, and are cultured in a selective medium (eg. hypoxanthine, aminopterin, thymidine medium, "HAT"). The resulting hybridomas are plated by limiting dilution, and are assayed for the production of antibodies which bind specifically to the immunizing antigen
- 25 30

(and which do not bind to unrelated antigens). The selected MAb-secreting hybridomas are then cultured either *in vitro* (eg. in tissue culture bottles or hollow fiber reactors), or *in vivo* (as ascites in mice).

If desired, the antibodies (whether polyclonal or monoclonal) may be labeled using conventional techniques. Suitable labels include fluorophores, chromophores, radioactive atoms (particularly ^{32}P and ^{125}I), electron-dense reagents, enzymes, and ligands having specific binding partners. Enzymes are typically detected by their activity. For example, horseradish peroxidase is usually detected by its ability to convert 3,3',5,5'-tetramethylbenzidine (TMB) to a blue pigment, quantifiable with a spectrophotometer. "Specific binding partner" refers to a protein capable of binding a ligand molecule with high specificity, as for example in the case of an antigen and a monoclonal antibody specific therefor. Other specific binding partners include biotin and avidin or streptavidin, IgG and protein A, and the numerous receptor-ligand couples known in the art. It should be understood that the above description is not meant to categorize the various labels into distinct classes, as the same label may serve in several different modes. For example, ^{125}I may serve as a radioactive label or as an electron-dense reagent. HRP may serve as enzyme or as antigen for a MAb. Further, one may combine various labels for desired effect. For example, MAbs and avidin also require labels in the practice of this invention: thus, one might label a MAb with biotin, and detect its presence with avidin labeled with ^{125}I , or with an anti-biotin MAb labeled with HRP. Other permutations and possibilities will be readily apparent to those of ordinary skill in the art, and are considered as equivalents within the scope of the instant invention.

Pharmaceutical Compositions

Pharmaceutical compositions can comprise either polypeptides, antibodies, or nucleic acid of the invention. The pharmaceutical compositions will comprise a therapeutically effective amount of either polypeptides, antibodies, or polynucleotides of the claimed invention.

The term "therapeutically effective amount" as used herein refers to an amount of a therapeutic agent to treat, ameliorate, or prevent a desired disease or condition, or to exhibit a detectable therapeutic or preventative effect. The effect can be detected by, for example, chemical markers or antigen levels. Therapeutic effects also include reduction in physical symptoms, such as decreased body temperature. The precise effective amount for a subject will depend upon the subject's size and health, the nature and extent of the condition, and the therapeutics or combination of

therapeutics selected for administration. Thus, it is not useful to specify an exact effective amount in advance. However, the effective amount for a given situation can be determined by routine experimentation and is within the judgement of the clinician.

For purposes of the present invention, an effective dose will be from about 0.01 mg/kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

A pharmaceutical composition can also contain a pharmaceutically acceptable carrier. The term "pharmaceutically acceptable carrier" refers to a carrier for administration of a therapeutic agent, such as antibodies or a polypeptide, genes, and other therapeutic agents. The term refers to any pharmaceutical carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition, and which may be administered without undue toxicity. Suitable carriers may be large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, and inactive virus particles. Such carriers are well known to those of ordinary skill in the art.

Pharmaceutically acceptable salts can be used therein, for example, mineral acid salts such as hydrochlorides, hydrobromides, phosphates, sulfates, and the like; and the salts of organic acids such as acetates, propionates, malonates, benzoates, and the like. A thorough discussion of pharmaceutically acceptable excipients is available in Remington's Pharmaceutical Sciences (Mack Pub. Co., N.J. 1991).

Pharmaceutically acceptable carriers in therapeutic compositions may contain liquids such as water, saline, glycerol and ethanol. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles. Typically, the therapeutic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. Liposomes are included within the definition of a pharmaceutically acceptable carrier.

25 Delivery Methods

Once formulated, the compositions of the invention can be administered directly to the subject. The subjects to be treated can be animals; in particular, human subjects can be treated.

Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal or transcutaneous applications (eg. see WO98/20734), needles, and gene guns or hyposprays. Dosage treatment may be a single dose schedule or a multiple dose schedule.

Vaccines

Vaccines according to the invention may either be prophylactic (ie. to prevent infection) or therapeutic (ie. to treat disease after infection).

Such vaccines comprise immunising antigen(s), immunogen(s), polypeptide(s), protein(s) or nucleic acid, usually in combination with "pharmaceutically acceptable carriers," which include any carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, lipid aggregates (such as oil droplets or liposomes), and inactive virus particles. Such carriers are well known to those of ordinary skill in the art. Additionally, these carriers may function as immunostimulating agents ("adjuvants"). Furthermore, the antigen or immunogen may be conjugated to a bacterial toxoid, such as a toxoid from diphtheria, tetanus, cholera, *H. pylori*, etc. pathogens.

Preferred adjuvants to enhance effectiveness of the composition include, but are not limited to: (1) aluminum salts (alum), such as aluminum hydroxide, aluminum phosphate, aluminum sulfate, etc; (2) oil-in-water emulsion formulations (with or without other specific immunostimulating agents such as muramyl peptides (see below) or bacterial cell wall components), such as for example (a) MF59™ (WO 90/14837; Chapter 10 in *Vaccine design: the subunit and adjuvant approach*, eds. Powell & Newman, Plenum Press 1995), containing 5% Squalene, 0.5% Tween 80, and 0.5% Span 85 (optionally containing various amounts of MTP-PE (see below), although not required) formulated into submicron particles using a microfluidizer such as Model 110Y microfluidizer (Microfluidics, Newton, MA), (b) SAF, containing 10% Squalene, 0.4% Tween 80, 5% pluronic-blocked polymer L121, and thr-MDP (see below) either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion, and (c) Ribi™ adjuvant system (RAS), (Ribi Immunochem, Hamilton, MT) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial

cell wall components from the group consisting of monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (Detox™); (3) saponin adjuvants, such as Stimulon™ (Cambridge Bioscience, Worcester, MA) may be used or particles generated therefrom such as ISCOMs (immunostimulating complexes); (4) Complete Freund's Adjuvant (CFA) and Incomplete Freund's Adjuvant (IFA); (5) cytokines, such as interleukins (*eg.* IL-1, IL-2, IL-4, IL-5, IL-6, IL-7, IL-12, *etc.*), interferons (*eg.* gamma interferon), macrophage colony stimulating factor (M-CSF), tumor necrosis factor (TNF), *etc.*; and (6) other substances that act as immunostimulating agents to enhance the effectiveness of the composition. Alum and MF59™ are preferred.

- 10 As mentioned above, muramyl peptides include, but are not limited to, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-normuramyl-L-alanyl-D-isoglutamine (nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-*sn*-glycero-3-hydroxyphosphoryloxy)-ethylamine (MTP-PE), *etc.*

- 15 The immunogenic compositions (*eg.* the immunising antigen/immunogen/polypeptide/protein/nucleic acid, pharmaceutically acceptable carrier, and adjuvant) typically will contain diluents, such as water, saline, glycerol, ethanol, *etc.* Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles.

- 20 Typically, the immunogenic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. The preparation also may be emulsified or encapsulated in liposomes for enhanced adjuvant effect, as discussed above under pharmaceutically acceptable carriers.

- 25 Immunogenic compositions used as vaccines comprise an immunologically effective amount of the antigenic or immunogenic polypeptides, as well as any other of the above-mentioned components, as needed. By "immunologically effective amount", it is meant that the administration of that amount to an individual, either in a single dose or as part of a series, is effective for treatment or prevention. This amount varies depending upon the health and physical condition of the individual to be treated, the taxonomic group of individual to be treated (*eg.* nonhuman primate, primate, *etc.*), the capacity of the individual's immune system to synthesize antibodies, the degree of protection desired, the formulation of the vaccine, the treating doctor's assessment of the medical situation,

and other relevant factors. It is expected that the amount will fall in a relatively broad range that can be determined through routine trials.

The immunogenic compositions are conventionally administered parenterally, *eg.* by injection, either subcutaneously, intramuscularly, or transdermally/transcutaneously (*eg.* WO98/20734).

5 Additional formulations suitable for other modes of administration include oral and pulmonary formulations, suppositories, and transdermal applications. Dosage treatment may be a single dose schedule or a multiple dose schedule. The vaccine may be administered in conjunction with other immunoregulatory agents.

As an alternative to protein-based vaccines, DNA vaccination may be employed [*eg.* Robinson & Torres (1997) *Seminars in Immunology* 9:271-283; Donnelly *et al.* (1997) *Annu Rev Immunol* 15:617-648; see later herein].

Gene Delivery Vehicles

Gene therapy vehicles for delivery of constructs including a coding sequence of a therapeutic of the invention, to be delivered to the mammal for expression in the mammal, can be administered either locally or systemically. These constructs can utilize viral or non-viral vector approaches in *in vivo* or *ex vivo* modality. Expression of such coding sequence can be induced using endogenous mammalian or heterologous promoters. Expression of the coding sequence *in vivo* can be either constitutive or regulated.

The invention includes gene delivery vehicles capable of expressing the contemplated nucleic acid sequences. The gene delivery vehicle is preferably a viral vector and, more preferably, a retroviral, adenoviral, adeno-associated viral (AAV), herpes viral, or alphavirus vector. The viral vector can also be an astrovirus, coronavirus, orthomyxovirus, papovavirus, paramyxovirus, parvovirus, picornavirus, poxvirus, or togavirus viral vector. See generally, Jolly (1994) *Cancer Gene Therapy* 1:51-64; Kimura (1994) *Human Gene Therapy* 5:845-852; Connelly (1995) *Human Gene Therapy* 6:185-193; and Kaplitt (1994) *Nature Genetics* 6:148-153.

Retroviral vectors are well known in the art and we contemplate that any retroviral gene therapy vector is employable in the invention, including B, C and D type retroviruses, xenotropic retroviruses (for example, NZB-X1, NZB-X2 and NZB9-1 (see O'Neill (1985) *J. Virol.* 53:160) polytropic retroviruses

eg. MCF and MCF-MLV (see Kelly (1983) *J. Virol.* 45:291), spumaviruses and lentiviruses. See RNA Tumor Viruses, Second Edition, Cold Spring Harbor Laboratory, 1985.

Portions of the retroviral gene therapy vector may be derived from different retroviruses. For example, retrovector LTRs may be derived from a Murine Sarcoma Virus, a tRNA binding site
5 from a Rous Sarcoma Virus, a packaging signal from a Murine Leukemia Virus, and an origin of second strand synthesis from an Avian Leukosis Virus.

These recombinant retroviral vectors may be used to generate transduction competent retroviral vector particles by introducing them into appropriate packaging cell lines (see US patent 5,591,624). Retrovirus vectors can be constructed for site-specific integration into host cell DNA
10 by incorporation of a chimeric integrase enzyme into the retroviral particle (see WO96/37626). It is preferable that the recombinant viral vector is a replication defective recombinant virus.

Packaging cell lines suitable for use with the above-described retrovirus vectors are well known in the art, are readily prepared (see WO95/30763 and WO92/05266), and can be used to create producer cell lines (also termed vector cell lines or "VCLs") for the production of recombinant
15 vector particles. Preferably, the packaging cell lines are made from human parent cells (eg. HT1080 cells) or mink parent cell lines, which eliminates inactivation in human serum.

Preferred retroviruses for the construction of retroviral gene therapy vectors include Avian Leukosis Virus, Bovine Leukemia Virus, Murine Leukemia Virus, Mink-Cell Focus-Inducing Virus, Murine Sarcoma Virus, Reticuloendotheliosis Virus and Rous Sarcoma Virus. Particularly
20 preferred Murine Leukemia Viruses include 4070A and 1504A (Hartley and Rowe (1976) *J Virol* 19:19-25), Abelson (ATCC No. VR-999), Friend (ATCC No. VR-245), Graffi, Gross (ATCC No. VR-590), Kirsten, Harvey Sarcoma Virus and Rauscher (ATCC No. VR-998) and Moloney Murine Leukemia Virus (ATCC No. VR-190). Such retroviruses may be obtained from depositories or collections such as the American Type Culture Collection ("ATCC") in Rockville, Maryland or
25 isolated from known sources using commonly available techniques.

Exemplary known retroviral gene therapy vectors employable in this invention include those described in patent applications GB2200651, EP0415731, EP0345242, EP0334301, WO89/02468; WO89/05349, WO89/09271, WO90/02806, WO90/07936, WO94/03622, WO93/25698,

WO93/25234, WO93/11230, WO93/10218, WO91/02805, WO91/02825, WO95/07994, US 5,219,740, US 4,405,712, US 4,861,719, US 4,980,289, US 4,777,127, US 5,591,624. See also Vile (1993) *Cancer Res* 53:3860-3864; Vile (1993) *Cancer Res* 53:962-967; Ram (1993) *Cancer Res* 53 (1993) 83-88; Takamiya (1992) *J Neurosci Res* 33:493-503; Baba (1993) *J Neurosurg* 79:729-735; Mann (1983) *Cell* 33:153; Cane (1984) *Proc Natl Acad Sci* 81:6349; and Miller (1990) *Human Gene Therapy* 1.

Human adenoviral gene therapy vectors are also known in the art and employable in this invention. See, for example, Berkner (1988) *Biotechniques* 6:616 and Rosenfeld (1991) *Science* 252:431, and WO93/07283, WO93/06223, and WO93/07282. Exemplary known adenoviral gene therapy vectors employable in this invention include those described in the above referenced documents and in WO94/12649, WO93/03769, WO93/19191, WO94/28938, WO95/11984, WO95/00655, WO95/27071, WO95/29993, WO95/34671, WO96/05320, WO94/08026, WO94/11506, WO93/06223, WO94/24299, WO95/14102, WO95/24297, WO95/02697, WO94/28152, WO94/24299, WO95/09241, WO95/25807, WO95/05835, WO94/18922 and WO95/09654.

Alternatively, administration of DNA linked to killed adenovirus as described in Curiel (1992) *Hum. Gene Ther.* 3:147-154 may be employed. The gene delivery vehicles of the invention also include adenovirus associated virus (AAV) vectors. Leading and preferred examples of such vectors for use in this invention are the AAV-2 based vectors disclosed in Srivastava, WO93/09239. Most preferred AAV vectors comprise the two AAV inverted terminal repeats in which the native D-sequences are modified by substitution of nucleotides, such that at least 5 native nucleotides and up to 18 native nucleotides, preferably at least 10 native nucleotides up to 18 native nucleotides, most preferably 10 native nucleotides are retained and the remaining nucleotides of the D-sequence are deleted or replaced with non-native nucleotides. The native D-sequences of the AAV inverted terminal repeats are sequences of 20 consecutive nucleotides in each AAV inverted terminal repeat (*ie.* there is one sequence at each end) which are not involved in HP formation. The non-native replacement nucleotide may be any nucleotide other than the nucleotide found in the native D-sequence in the same position. Other employable exemplary AAV vectors are pWP-19, pWN-1, both of which are disclosed in Nahreini (1993) *Gene* 124:257-262. Another example of such an AAV vector is psub201 (see Samulski (1987) *J. Virol.* 61:3096). Another exemplary AAV vector is the Double-D ITR vector. Construction of the Double-D ITR vector is disclosed in US Patent 5,478,745. Still other vectors are those disclosed in Carter US Patent 4,797,368 and

Muzyczka US Patent 5,139,941, Chartejee US Patent 5,474,935, and Kotin WO94/288157. Yet a further example of an AAV vector employable in this invention is SSV9AFABTKneo, which contains the AFP enhancer and albumin promoter and directs expression predominantly in the liver. Its structure and construction are disclosed in Su (1996) *Human Gene Therapy* 7:463-470.

- 5 Additional AAV gene therapy vectors are described in US 5,354,678, US 5,173,414, US 5,139,941, and US 5,252,479.

The gene therapy vectors of the invention also include herpes vectors. Leading and preferred examples are herpes simplex virus vectors containing a sequence encoding a thymidine kinase polypeptide such as those disclosed in US 5,288,641 and EP0176170 (Roizman). Additional
10 exemplary herpes simplex virus vectors include HFEM/ICP6-LacZ disclosed in WO95/04139 (Wistar Institute), pHSVlac described in Geller (1988) *Science* 241:1667-1669 and in WO90/09441 and WO92/07945, HSV Us3::pgC-lacZ described in Fink (1992) *Human Gene Therapy* 3:11-19 and HSV 7134, 2 RH 105 and GAL4 described in EP 0453242 (Breakefield), and those deposited with the ATCC as accession numbers ATCC VR-977 and ATCC VR-260.

- 15 Also contemplated are alpha virus gene therapy vectors that can be employed in this invention. Preferred alpha virus vectors are Sindbis viruses vectors. Togaviruses, Semliki Forest virus (ATCC VR-67; ATCC VR-1247), Middleberg virus (ATCC VR-370), Ross River virus (ATCC VR-373; ATCC VR-1246), Venezuelan equine encephalitis virus (ATCC VR923; ATCC VR-1250; ATCC VR-1249; ATCC VR-532), and those described in US patents 5,091,309, 5,217,879, and
20 WO92/10578. More particularly, those alpha virus vectors described in US Serial No. 08/405,627, filed March 15, 1995, WO94/21792, WO92/10578, WO95/07994, US 5,091,309 and US 5,217,879 are employable. Such alpha viruses may be obtained from depositories or collections such as the ATCC in Rockville, Maryland or isolated from known sources using commonly available techniques. Preferably, alphavirus vectors with reduced cytotoxicity are used (see USSN
25 08/679640).

DNA vector systems such as eukaryotic layered expression systems are also useful for expressing the nucleic acids of the invention. See WO95/07994 for a detailed description of eukaryotic layered expression systems. Preferably, the eukaryotic layered expression systems of the invention are derived from alphavirus vectors and most preferably from Sindbis viral vectors.

Other viral vectors suitable for use in the present invention include those derived from poliovirus, for example ATCC VR-58 and those described in Evans, *Nature* 339 (1989) 385 and Sabin (1973) *J. Biol. Standardization* 1:115; rhinovirus, for example ATCC VR-1110 and those described in Arnold (1990) *J Cell Biochem* 40:1; pox viruses such as canary pox virus or vaccinia virus, for example ATCC VR-111 and ATCC VR-2010 and those described in Fisher-Hoch (1989) *Proc Natl Acad Sci* 86:317; Flexner (1989) *Ann NY Acad Sci* 569:86, Flexner (1990) *Vaccine* 8:17; in US 4,603,112 and US 4,769,330 and WO89/01973; SV40 virus, for example ATCC VR-305 and those described in Mulligan (1979) *Nature* 277:108 and Madzak (1992) *J Gen Virol* 73:1533; influenza virus, for example ATCC VR-797 and recombinant influenza viruses made employing reverse genetics techniques as described in US 5,166,057 and in Enami (1990) *Proc Natl Acad Sci* 87:3802-3805; Enami & Palese (1991) *J Virol* 65:2711-2713 and Luytjes (1989) *Cell* 59:110, (see also McMichael (1983) *NEJ Med* 309:13, and Yap (1978) *Nature* 273:238 and *Nature* (1979) 277:108); human immunodeficiency virus as described in EP-0386882 and in Buchschacher (1992) *J. Virol.* 66:2731; measles virus, for example ATCC VR-67 and VR-1247 and those described in EP-0440219; Aura virus, for example ATCC VR-368; Bebaru virus, for example ATCC VR-600 and ATCC VR-1240; Cabassou virus, for example ATCC VR-922; Chikungunya virus, for example ATCC VR-64 and ATCC VR-1241; Fort Morgan Virus, for example ATCC VR-924; Getah virus, for example ATCC VR-369 and ATCC VR-1243; Kyzylagach virus, for example ATCC VR-927; Mayaro virus, for example ATCC VR-66; Mucambo virus, for example ATCC VR-580 and ATCC VR-1244; Ndumu virus, for example ATCC VR-371; Pixuna virus, for example ATCC VR-372 and ATCC VR-1245; Tonate virus, for example ATCC VR-925; Trinita virus, for example ATCC VR-469; Una virus, for example ATCC VR-374; Whataroa virus, for example ATCC VR-926; Y-62-33 virus, for example ATCC VR-375; O'Nyong virus, Eastern encephalitis virus, for example ATCC VR-65 and ATCC VR-1242; Western encephalitis virus, for example ATCC VR-70, ATCC VR-1251, ATCC VR-622 and ATCC VR-1252; and coronavirus, for example ATCC VR-740 and those described in Hamre (1966) *Proc Soc Exp Biol Med* 121:190.

Delivery of the compositions of this invention into cells is not limited to the above mentioned viral vectors. Other delivery methods and media may be employed such as, for example, nucleic acid expression vectors, polycationic condensed DNA linked or unlinked to killed adenovirus alone, for example see US Serial No. 08/366,787, filed December 30, 1994 and Curiel (1992) *Hum Gene Ther* 3:147-154 ligand linked DNA, for example see Wu (1989) *J Biol Chem* 264:16985-16987, eucaryotic cell delivery vehicles cells, for example see US Serial No.08/240,030, filed May 9,

1994, and US Serial No. 08/404,796, deposition of photopolymerized hydrogel materials, hand-held gene transfer particle gun, as described in US Patent 5,149,655, ionizing radiation as described in US5,206,152 and in WO92/11033, nucleic charge neutralization or fusion with cell membranes. Additional approaches are described in Philip (1994) *Mol Cell Biol* 14:2411-2418 and
5 in Woffendin (1994) *Proc Natl Acad Sci* 91:1581-1585.

Particle mediated gene transfer may be employed, for example see US Serial No. 60/023,867. Briefly, the sequence can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell targeting
10 ligands such as asialoorosomucoid, as described in Wu & Wu (1987) *J. Biol. Chem.* 262:4429-4432, insulin as described in Hucked (1990) *Biochem Pharmacol* 40:253-263, galactose as described in Plank (1992) *Bioconjugate Chem* 3:533-539, lactose or transferrin.

Naked DNA may also be employed. Exemplary naked DNA introduction methods are described in WO 90/11092 and US 5,580,859. Uptake efficiency may be improved using biodegradable latex
15 beads. DNA coated latex beads are efficiently transported into cells after endocytosis initiation by the beads. The method may be improved further by treatment of the beads to increase hydrophobicity and thereby facilitate disruption of the endosome and release of the DNA into the cytoplasm.

Liposomes that can act as gene delivery vehicles are described in US 5,422,120, WO95/13796, WO94/23697, WO91/14445 and EP-524,968. As described in USSN. 60/023,867, on non-viral
20 delivery, the nucleic acid sequences encoding a polypeptide can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then be incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, insulin, galactose, lactose, or transferrin. Other delivery systems include the use of liposomes to encapsulate
25 DNA comprising the gene under the control of a variety of tissue-specific or ubiquitously-active promoters. Further non-viral delivery suitable for use includes mechanical delivery systems such as the approach described in Woffendin *et al* (1994) *Proc. Natl. Acad. Sci. USA* 91(24):11581-11585. Moreover, the coding sequence and the product of expression of such can be delivered through deposition of photopolymerized hydrogel materials. Other conventional methods
30 for gene delivery that can be used for delivery of the coding sequence include, for example, use of

hand-held gene transfer particle gun, as described in US 5,149,655; use of ionizing radiation for activating transferred gene, as described in US 5,206,152 and WO92/11033

Exemplary liposome and polycationic gene delivery vehicles are those described in US 5,422,120 and 4,762,915; in WO 95/13796; WO94/23697; and WO91/14445; in EP-0524968; and in Stryer,

- 5 Biochemistry, pages 236-240 (1975) W.H. Freeman, San Francisco; Szoka (1980) *Biochem Biophys Acta* 600:1; Bayer (1979) *Biochem Biophys Acta* 550:464; Rivnay (1987) *Meth Enzymol* 149:119; Wang (1987) *Proc Natl Acad Sci* 84:7851; Plant (1989) *Anal Biochem* 176:420.

- A polynucleotide composition can comprises therapeutically effective amount of a gene therapy vehicle, as the term is defined above. For purposes of the present invention, an effective dose will
10 be from about 0.01 mg/ kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

Delivery Methods

- Once formulated, the polynucleotide compositions of the invention can be administered (1) directly to the subject; (2) delivered *ex vivo*, to cells derived from the subject; or (3) *in vitro* for expression
15 of recombinant proteins. The subjects to be treated can be mammals or birds. Also, human subjects can be treated.

- Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of
20 administration include oral and pulmonary administration, suppositories, and transdermal or transcutaneous applications (eg. see WO98/20734), needles, and gene guns or hypodermic sprays. Dosage treatment may be a single dose schedule or a multiple dose schedule.

- Methods for the *ex vivo* delivery and reimplantation of transformed cells into a subject are known in the art and described in eg. WO93/14778. Examples of cells useful in *ex vivo* applications
25 include, for example, stem cells, particularly hematopoietic, lymph cells, macrophages, dendritic cells, or tumor cells.

Generally, delivery of nucleic acids for both *ex vivo* and *in vitro* applications can be accomplished by the following procedures, for example, dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei, all well known in the art.

Polynucleotide and polypeptide pharmaceutical compositions

In addition to the pharmaceutically acceptable carriers and salts described above, the following additional agents can be used with polynucleotide and/or polypeptide compositions.

A. Polypeptides

- One example are polypeptides which include, without limitation: asialoglycoprotein (ASOR); transferrin; asialoglycoproteins; antibodies; antibody fragments; ferritin; interleukins; interferons, granulocyte, macrophage colony stimulating factor (GM-CSF), granulocyte colony stimulating factor (G-CSF), macrophage colony stimulating factor (M-CSF), stem cell factor and erythropoietin. Viral antigens, such as envelope proteins, can also be used. Also, proteins from other invasive organisms, such as the 17 amino acid peptide from the circumsporozoite protein of plasmodium falciparum known as RII.

B. Hormones, Vitamins, etc.

Other groups that can be included are, for example: hormones, steroids, androgens, estrogens, thyroid hormone, or vitamins, folic acid.

C. Polyalkylenes, Polysaccharides, etc.

Also, polyalkylene glycol can be included with the desired polynucleotides/polypeptides. In a preferred embodiment, the polyalkylene glycol is polyethylene glycol. In addition, mono-, di-, or polysaccharides can be included. In a preferred embodiment of this aspect, the polysaccharide is dextran or DEAE-dextran. Also, chitosan and poly(lactide-co-glycolide)

D. Lipids, and Liposomes

The desired polynucleotide/polypeptide can also be encapsulated in lipids or packaged in liposomes prior to delivery to the subject or to cells derived therefrom.

Lipid encapsulation is generally accomplished using liposomes which are able to stably bind or entrap and retain nucleic acid. The ratio of condensed polynucleotide to lipid preparation can vary but will generally be around 1:1 (mg DNA:micromoles lipid), or more of lipid. For a review of the use of liposomes as carriers for delivery of nucleic acids, see, Hug and Sleight (1991) *Biochim. Biophys. Acta.* 1097:1-17; Straubinger (1983) *Meth. Enzymol.* 101:512-527.

Liposomal preparations for use in the present invention include cationic (positively charged), anionic (negatively charged) and neutral preparations. Cationic liposomes have been shown to mediate intracellular delivery of plasmid DNA (Felgner (1987) *Proc. Natl. Acad. Sci. USA* 84:7413-7416); mRNA (Malone (1989) *Proc. Natl. Acad. Sci. USA* 86:6077-6081); and purified transcription factors (Debs (1990) *J. Biol. Chem.* 265:10189-10192), in functional form.

Cationic liposomes are readily available. For example, N[1-2,3-dioleoyloxy)propyl]-N,N,N-triethylammonium (DOTMA) liposomes are available under the trademark Lipofectin, from GIBCO BRL, Grand Island, NY. (See, also, Felgner *supra*). Other commercially available liposomes include transfectace (DDAB/DOPE) and DOTAP/DOPE (Boehringer). Other cationic liposomes can be prepared from readily available materials using techniques well known in the art. See, *eg.* Szoka (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; WO90/11092 for a description of the synthesis of DOTAP (1,2-bis(oleoyloxy)-3-(trimethylammonio)propane) liposomes.

Similarly, anionic and neutral liposomes are readily available, such as from Avanti Polar Lipids (Birmingham, AL), or can be easily prepared using readily available materials. Such materials include phosphatidyl choline, cholesterol, phosphatidyl ethanolamine, dioleoylphosphatidyl choline (DOPC), dioleoylphosphatidyl glycerol (DOPG), dioleoylphosphatidyl ethanolamine (DOPE), among others. These materials can also be mixed with the DOTMA and DOTAP starting materials in appropriate ratios. Methods for making liposomes using these materials are well known in the art.

The liposomes can comprise multilamellar vesicles (MLVs), small unilamellar vesicles (SUVs), or large unilamellar vesicles (LUVs). The various liposome-nucleic acid complexes are prepared using methods known in the art. See *eg.* Straubinger (1983) *Meth. Immunol.* 101:512-527; Szoka (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; Papahadjopoulos (1975) *Biochim. Biophys. Acta* 394:483; Wilson (1979) *Cell* 17:77; Deamer & Bangham (1976) *Biochim. Biophys. Acta* 443:629; Ostro (1977) *Biochem. Biophys. Res. Commun.* 76:836; Fraley (1979) *Proc. Natl. Acad. Sci. USA*

76:3348); Enoch & Strittmatter (1979) *Proc. Natl. Acad. Sci. USA* 76:145; Fraley (1980) *J. Biol. Chem.* (1980) 255:10431; Szoka & Papahadjopoulos (1978) *Proc. Natl. Acad. Sci. USA* 75:145; and Schaefer-Ridder (1982) *Science* 215:166.

E.Lipoproteins

- 5 In addition, lipoproteins can be included with the polynucleotide/polypeptide to be delivered. Examples of lipoproteins to be utilized include: chylomicrons, HDL, IDL, LDL, and VLDL. Mutants, fragments, or fusions of these proteins can also be used. Also, modifications of naturally occurring lipoproteins can be used, such as acetylated LDL. These lipoproteins can target the delivery of polynucleotides to cells expressing lipoprotein receptors. Preferably, if lipoproteins are including with
10 the polynucleotide to be delivered, no other targeting ligand is included in the composition.

Naturally occurring lipoproteins comprise a lipid and a protein portion. The protein portion are known as apoproteins. At the present, apoproteins A, B, C, D, and E have been isolated and identified. At least two of these contain several proteins, designated by Roman numerals, AI, AII, AIV; CI, CII, CIII.

- 15 A lipoprotein can comprise more than one apoprotein. For example, naturally occurring chylomicrons comprises of A, B, C, and E, over time these lipoproteins lose A and acquire C and E apoproteins. VLDL comprises A, B, C, and E apoproteins, LDL comprises apoprotein B; and HDL comprises apoproteins A, C, and E.

- The amino acid of these apoproteins are known and are described in, for example, Breslow (1985)
20 *Annu Rev. Biochem* 54:699; Law (1986) *Adv. Exp Med. Biol.* 151:162; Chen (1986) *J Biol Chem* 261:12918; Kane (1980) *Proc Natl Acad Sci USA* 77:2465; and Utermann (1984) *Hum Genet* 65:232.

- Lipoproteins contain a variety of lipids including, triglycerides, cholesterol (free and esters), and phospholipids. The composition of the lipids varies in naturally occurring lipoproteins. For example, chylomicrons comprise mainly triglycerides. A more detailed description of the lipid
25 content of naturally occurring lipoproteins can be found, for example, in *Meth. Enzymol.* 128 (1986). The composition of the lipids are chosen to aid in conformation of the apoprotein for receptor binding activity. The composition of lipids can also be chosen to facilitate hydrophobic interaction and association with the polynucleotide binding molecule.

Naturally occurring lipoproteins can be isolated from serum by ultracentrifugation, for instance. Such methods are described in *Meth. Enzymol.* (*supra*); Pitas (1980) *J. Biochem.* 255:5454-5460 and Mahey (1979) *J Clin. Invest* 64:743-750. Lipoproteins can also be produced by *in vitro* or recombinant methods by expression of the apoprotein genes in a desired host cell. See, for example, 5 Atkinson (1986) *Annu Rev Biophys Chem* 15:403 and Radding (1958) *Biochim Biophys Acta* 30: 443. Lipoproteins can also be purchased from commercial suppliers, such as Biomedical Technologies, Inc., Stoughton, Massachusetts, USA. Further description of lipoproteins can be found in Zuckermann *et al.* PCT/US97/14465.

F. Polycationic Agents

- 10 Polycationic agents can be included, with or without lipoprotein, in a composition with the desired polynucleotide/polypeptide to be delivered.

Polycationic agents, typically, exhibit a net positive charge at physiological relevant pH and are capable of neutralizing the electrical charge of nucleic acids to facilitate delivery to a desired location. These agents have both *in vitro*, *ex vivo*, and *in vivo* applications. Polycationic agents can 15 be used to deliver nucleic acids to a living subject either intramuscularly, subcutaneously, etc.

The following are examples of useful polypeptides as polycationic agents: polylysine, polyarginine, polyornithine, and protamine. Other examples include histones, protamines, human serum albumin, DNA binding proteins, non-histone chromosomal proteins, coat proteins from DNA viruses, such as (X174, transcriptional factors also contain domains that bind DNA and therefore may be useful 20 as nucleic acid condensing agents. Briefly, transcriptional factors such as C/CEBP, c-jun, c-fos, AP-1, AP-2, AP-3, CPF, Prot-1, Sp-1, Oct-1, Oct-2, CREP, and TFIID contain basic domains that bind DNA sequences.

Organic polycationic agents include: spermine, spermidine, and putrescine.

The dimensions and of the physical properties of a polycationic agent can be extrapolated from the 25 list above, to construct other polypeptide polycationic agents or to produce synthetic polycationic agents.

Synthetic polycationic agents which are useful include, for example, DEAE-dextran, polybrene. Lipofectin™, and lipofectAMINE™ are monomers that form polycationic complexes when combined with polynucleotides/polypeptides.

Immunodiagnostic Assays

5 Neisserial antigens of the invention can be used in immunoassays to detect antibody levels (or, conversely, anti-Neisserial antibodies can be used to detect antigen levels). Immunoassays based on well defined, recombinant antigens can be developed to replace invasive diagnostics methods. Antibodies to Neisserial proteins within biological samples, including for example, blood or serum samples, can be detected. Design of the immunoassays is subject to a great deal of variation, and
10 a variety of these are known in the art. Protocols for the immunoassay may be based, for example, upon competition, or direct reaction, or sandwich type assays. Protocols may also, for example, use solid supports, or may be by immunoprecipitation. Most assays involve the use of labeled antibody or polypeptide; the labels may be, for example, fluorescent, chemiluminescent, radioactive, or dye molecules. Assays which amplify the signals from the probe are also known; examples of which
15 are assays which utilize biotin and avidin, and enzyme-labeled and mediated immunoassays, such as ELISA assays.

Kits suitable for immunodiagnosis and containing the appropriate labeled reagents are constructed by packaging the appropriate materials, including the compositions of the invention, in suitable containers, along with the remaining reagents and materials (for example, suitable buffers, salt
20 solutions, etc.) required for the conduct of the assay, as well as suitable set of assay instructions.

Nucleic Acid Hybridisation

"Hybridization" refers to the association of two nucleic acid sequences to one another by hydrogen bonding. Typically, one sequence will be fixed to a solid support and the other will be free in solution. Then, the two sequences will be placed in contact with one another under conditions that favor
25 hydrogen bonding. Factors that affect this bonding include: the type and volume of solvent; reaction temperature; time of hybridization; agitation; agents to block the non-specific attachment of the liquid phase sequence to the solid support (Denhardt's reagent or BLOTTO); concentration of the sequences; use of compounds to increase the rate of association of sequences (dextran sulfate or polyethylene glycol); and the stringency of the washing conditions following hybridization. See Sambrook *et al.*
30 [*supra*] Volume 2, chapter 9, pages 9.47 to 9.57.

"Stringency" refers to conditions in a hybridization reaction that favor association of very similar sequences over sequences that differ. For example, the combination of temperature and salt concentration should be chosen that is approximately 120 to 200°C below the calculated T_m of the hybrid under study. The temperature and salt conditions can often be determined empirically in preliminary experiments in which samples of genomic DNA immobilized on filters are hybridized to the sequence of interest and then washed under conditions of different stringencies. See Sambrook *et al.* at page 9.50.

Variables to consider when performing, for example, a Southern blot are (1) the complexity of the DNA being blotted and (2) the homology between the probe and the sequences being detected. The total amount of the fragment(s) to be studied can vary a magnitude of 10, from 0.1 to 1 µg for a plasmid or phage digest to 10^{-9} to 10^{-8} g for a single copy gene in a highly complex eukaryotic genome. For lower complexity polynucleotides, substantially shorter blotting, hybridization, and exposure times, a smaller amount of starting polynucleotides, and lower specific activity of probes can be used. For example, a single-copy yeast gene can be detected with an exposure time of only 1 hour starting with 1 µg of yeast DNA, blotting for two hours, and hybridizing for 4-8 hours with a probe of 10^8 cpm/µg. For a single-copy mammalian gene a conservative approach would start with 10 µg of DNA, blot overnight, and hybridize overnight in the presence of 10% dextran sulfate using a probe of greater than 10^8 cpm/µg, resulting in an exposure time of ~24 hours.

Several factors can affect the melting temperature (T_m) of a DNA-DNA hybrid between the probe and the fragment of interest, and consequently, the appropriate conditions for hybridization and washing. In many cases the probe is not 100% homologous to the fragment. Other commonly encountered variables include the length and total G+C content of the hybridizing sequences and the ionic strength and formamide content of the hybridization buffer. The effects of all of these factors can be approximated by a single equation:

$$T_m = 81 + 16.6(\log_{10} C_i) + 0.4[\%(G + C)] - 0.6(\%\text{formamide}) - 600/n - 1.5(\%\text{mismatch}).$$

where C_i is the salt concentration (monovalent ions) and n is the length of the hybrid in base pairs (slightly modified from Meinkoth & Wahl (1984) *Anal. Biochem.* 138: 267-284).

In designing a hybridization experiment, some factors affecting nucleic acid hybridization can be conveniently altered. The temperature of the hybridization and washes and the salt concentration during the washes are the simplest to adjust. As the temperature of the hybridization increases (*ie.* stringency), it becomes less likely for hybridization to occur between strands that are nonhomologous, and as a result, background decreases. If the radiolabeled probe is not completely homologous with the immobilized fragment (as is frequently the case in gene family and interspecies hybridization experiments), the hybridization temperature must be reduced, and background will increase. The temperature of the washes affects the intensity of the hybridizing band and the degree of background in a similar manner. The stringency of the washes is also increased with decreasing salt concentrations.

In general, convenient hybridization temperatures in the presence of 50% formamide are 42°C for a probe with is 95% to 100% homologous to the target fragment, 37°C for 90% to 95% homology, and 32°C for 85% to 90% homology. For lower homologies, formamide content should be lowered and temperature adjusted accordingly, using the equation above. If the homology between the probe and the target fragment are not known, the simplest approach is to start with both hybridization and wash conditions which are nonstringent. If non-specific bands or high background are observed after autoradiography, the filter can be washed at high stringency and reexposed. If the time required for exposure makes this approach impractical, several hybridization and/or washing stringencies should be tested in parallel.

Nucleic Acid Probe Assays

Methods such as PCR, branched DNA probe assays, or blotting techniques utilizing nucleic acid probes according to the invention can determine the presence of cDNA or mRNA. A probe is said to "hybridize" with a sequence of the invention if it can form a duplex or double stranded complex, which is stable enough to be detected.

The nucleic acid probes will hybridize to the Neisserial nucleotide sequences of the invention (including both sense and antisense strands). Though many different nucleotide sequences will encode the amino acid sequence, the native Neisserial sequence is preferred because it is the actual sequence present in cells. mRNA represents a coding sequence and so a probe should be complementary to the coding sequence; single-stranded cDNA is complementary to mRNA, and so a cDNA probe should be complementary to the non-coding sequence.

The probe sequence need not be identical to the Neisserial sequence (or its complement) — some variation in the sequence and length can lead to increased assay sensitivity if the nucleic acid probe can form a duplex with target nucleotides, which can be detected. Also, the nucleic acid probe can include additional nucleotides to stabilize the formed duplex. Additional Neisserial sequence may
5 also be helpful as a label to detect the formed duplex. For example, a non-complementary nucleotide sequence may be attached to the 5' end of the probe, with the remainder of the probe sequence being complementary to a Neisserial sequence. Alternatively, non-complementary bases or longer sequences can be interspersed into the probe, provided that the probe sequence has sufficient complementarity with the a Neisserial sequence in order to hybridize therewith and
10 thereby form a duplex which can be detected.

The exact length and sequence of the probe will depend on the hybridization conditions, such as temperature, salt condition and the like. For example, for diagnostic applications, depending on the complexity of the analyte sequence, the nucleic acid probe typically contains at least 10-20 nucleotides, preferably 15-25, and more preferably at least 30 nucleotides, although it may be
15 shorter than this. Short primers generally require cooler temperatures to form sufficiently stable hybrid complexes with the template.

Probes may be produced by synthetic procedures, such as the triester method of Matteucci *et al.* [*J. Am. Chem. Soc.* (1981) 103:3185], or according to Urdea *et al.* [*Proc. Natl. Acad. Sci. USA* (1983) 80: 7461], or using commercially available automated oligonucleotide synthesizers.

20 The chemical nature of the probe can be selected according to preference. For certain applications, DNA or RNA are appropriate. For other applications, modifications may be incorporated *eg.* backbone modifications, such as phosphorothioates or methylphosphonates, can be used to increase *in vivo* half-life, alter RNA affinity, increase nuclease resistance *etc.* [*eg.* see Agrawal & Iyer (1995) *Curr Opin Biotechnol* 6:12-19; Agrawal (1996) *TIBTECH* 14:376-387]; analogues such as
25 peptide nucleic acids may also be used [*eg.* see Corey (1997) *TIBTECH* 15:224-229; Buchardt *et al.* (1993) *TIBTECH* 11:384-386].

Alternatively, the polymerase chain reaction (PCR) is another well-known means for detecting small amounts of target nucleic acids. The assay is described in: Mullis *et al.* [*Meth. Enzymol.* (1987) 155: 335-350]; US patents 4,683,195 and 4,683,202. Two "primer" nucleotides hybridize

with the target nucleic acids and are used to prime the reaction. The primers can comprise sequence that does not hybridize to the sequence of the amplification target (or its complement) to aid with duplex stability or, for example, to incorporate a convenient restriction site. Typically, such sequence will flank the desired Neisserial sequence.

- 5 A thermostable polymerase creates copies of target nucleic acids from the primers using the original target nucleic acids as a template. After a threshold amount of target nucleic acids are generated by the polymerase, they can be detected by more traditional methods, such as Southern blots. When using the Southern blot method, the labelled probe will hybridize to the Neisserial sequence (or its complement).
- 10 Also, mRNA or cDNA can be detected by traditional blotting techniques described in Sambrook *et al* [*supra*]. mRNA, or cDNA generated from mRNA using a polymerase enzyme, can be purified and separated using gel electrophoresis. The nucleic acids on the gel are then blotted onto a solid support, such as nitrocellulose. The solid support is exposed to a labelled probe and then washed to remove any unhybridized probe. Next, the duplexes containing the labeled probe are detected.
- 15 Typically, the probe is labelled with a radioactive moiety.

BRIEF DESCRIPTION OF THE DRAWINGS

- Figures 1-7 show biochemical data and sequence analysis pertaining to Examples 1, 2, 3, 7, 13, 16 and 19, respectively, with ORFs 40, 38, 44, 52, 114, 41 and 124.. M1 and M2 are molecular weight markers. Arrows indicate the position of the main recombinant product or, in Western blots, the position of the main *N.meningitidis* immunoreactive band. TP indicates *N.meningitidis* total protein extract; OMV indicates *N.meningitidis* outer membrane vesicle preparation. In bactericidal assay results: a diamond (♦) shows preimmune data; a triangle (▲) shows GST control data; a circle (●) shows data with recombinant *N.meningitidis* protein. Computer analyses show a hydrophilicity plot (upper), an antigenic index plot (middle), and an AMPHI analysis (lower). The
- 20 AMPHI program has been used to predict T-cell epitopes [Gao *et al.* (1989) *J. Immunol.* 143:3007; Roberts *et al.* (1996) *AIDS Res Hum Retrovir* 12:593; Quakyi *et al.* (1992) *Scand J Immunol* suppl.11:9) and is available in the Protean package of DNASTAR, Inc. (1228 South Park Street, Madison, Wisconsin 53715 USA).
- 25

EXAMPLES

The examples describe nucleic acid sequences which have been identified in *N.meningitidis*, along with their putative translation products. Not all of the nucleic acid sequences are complete *ie.* they encode less than the full-length wild-type protein. It is believed at present that none of the DNA sequences described herein have significant homologs in *N.gonorrhoeae*.

The examples are generally in the following format:

- a nucleotide sequence which has been identified in *N.meningitidis* (strain B)
- the putative translation product of this sequence
- a computer analysis of the translation product based on database comparisons
- a corresponding gene and protein sequence identified in *N.meningitidis* (strain A)
- a description of the characteristics of the proteins which indicates that they might be suitably antigenic
- results of biochemical analysis (expression, purification, ELISA, FACS *etc.*)

The examples typically include details of sequence homology between species and strains. Proteins that are similar in sequence are generally similar in both structure and function, and the homology often indicates a common evolutionary origin. Comparison with sequences of proteins of known function is widely used as a guide for the assignment of putative protein function to a new sequence and has proved particularly useful in whole-genome analyses.

Sequence comparisons were performed at NCBI (<http://www.ncbi.nlm.nih.gov>) using the algorithms BLAST, BLAST2, BLASTn, BLASTp, tBLASTn, BLASTx, & tBLASTx [*eg.* see also Altschul *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25:2289-3402]. Searches were performed against the following databases: non-redundant GenBank+EMBL+DDBJ+PDB sequences and non-redundant GenBank CDS translations+PDB+SwissProt+SPupdate+PIR sequences.

Dots within nucleotide sequences (*eg.* position 288 in Example 12) represent nucleotides which have been arbitrarily introduced in order to maintain a reading frame. In the same way, double-underlined nucleotides were removed. Lower case letters (*eg.* position 589 in Example 12) represent ambiguities which arose during alignment of independent sequencing reactions (some of

the nucleotide sequences in the examples are derived from combining the results of two or more experiments).

Nucleotide sequences were scanned in all six reading frames to predict the presence of hydrophobic domains using an algorithm based on the statistical studies of Esposti *et al.* [Critical evaluation of the hydropathy of membrane proteins (1990) *Eur J Biochem* 190:207-219]. These domains
5 represent potential transmembrane regions or hydrophobic leader sequences.

Open reading frames were predicted from fragmented nucleotide sequences using the program ORFFINDER (NCBI).

Underlined amino acid sequences indicate possible transmembrane domains or leader sequences
10 in the ORFs, as predicted by the PSORT algorithm (<http://www.psort.nibb.ac.jp>). Functional domains were also predicted using the MOTIFS program (GCG Wisconsin & PROSITE).

Various tests can be used to assess the *in vivo* immunogenicity of the proteins identified in the examples. For example, the proteins can be expressed recombinantly and used to screen patient sera by immunoblot. A positive reaction between the protein and patient serum indicates that the patient
15 has previously mounted an immune response to the protein in question *ie.* the protein is an immunogen. This method can also be used to identify immunodominant proteins.

The recombinant protein can also be conveniently used to prepare antibodies *eg.* in a mouse. These can be used for direct confirmation that a protein is located on the cell-surface. Labelled antibody (*eg.* fluorescent labelling for FACS) can be incubated with intact bacteria and the presence of label
20 on the bacterial surface confirms the location of the protein.

In particular, the following methods (A) to (S) were used to express, purify and biochemically characterise the proteins of the invention:

A) Chromosomal DNA preparation

N.meningitidis strain 2996 was grown to exponential phase in 100ml of GC medium, harvested by
25 centrifugation, and resuspended in 5ml buffer (20% Sucrose, 50mM Tris-HCl, 50mM EDTA, pH8). After 10 minutes incubation on ice, the bacteria were lysed by adding 10ml lysis solution (50mM NaCl, 1% Na-Sarkosyl, 50µg/ml Proteinase K), and the suspension was incubated at 37°C for 2

hours. Two phenol extractions (equilibrated to pH 8) and one ChCl_3 /isoamylalcohol (24:1) extraction were performed. DNA was precipitated by addition of 0.3M sodium acetate and 2 volumes ethanol, and was collected by centrifugation. The pellet was washed once with 70% ethanol and redissolved in 4ml buffer (10mM Tris-HCl, 1mM EDTA, pH 8). The DNA concentration was measured by reading the OD at 260 nm.

B) Oligonucleotide design

Synthetic oligonucleotide primers were designed on the basis of the coding sequence of each ORF, using (a) the meningococcus B sequence when available, or (b) the gonococcus/meningococcus A sequence, adapted to the codon preference usage of meningococcus as necessary. Any predicted signal peptides were omitted, by deducing the 5'-end amplification primer sequence immediately downstream from the predicted leader sequence.

The 5' primers included two restriction enzyme recognition sites (*Bam*HI-*Nde*I, *Bam*HI-*Nhe*I, or *Eco*RI-*Nhe*I, depending on the gene's own restriction pattern); the 3' primers included a *Xho*I restriction site. This procedure was established in order to direct the cloning of each amplification product (corresponding to each ORF) into two different expression systems: pGEX-KG (using either *Bam*HI-*Xho*I or *Eco*RI-*Xho*I), and pET21b+ (using either *Nde*I-*Xho*I or *Nhe*I-*Xho*I).

5'-end primer tail: CGCGGATCCCATATG (*Bam*HI-*Nde*I)

CGCGGATCCGCTAGC (*Bam*HI-*Nhe*I)

CCGGAATTCTAGCTAGC (*Eco*RI-*Nhe*I)

3'-end primer tail: CCCGCTCGAG (*Xho*I)

As well as containing the restriction enzyme recognition sequences, the primers included nucleotides which hybridised to the sequence to be amplified. The number of hybridizing nucleotides depended on the melting temperature of the whole primer, and was determined for each primer using the formulae:

$$T_m = 4 (G+C) + 2 (A+T) \quad (\text{tail excluded})$$

$$T_m = 64.9 + 0.41 (\% \text{ GC}) - 600/N \quad (\text{whole primer})$$

The average melting temperature of the selected oligos were 65-70°C for the whole oligo and 50-55°C for the hybridising region alone.

Table I shows the forward and reverse primers used for each amplification. Oligos were synthesized by a Perkin Elmer 394 DNA/RNA Synthesizer, eluted from the columns in 2ml NH_4OH , and deprotected by 5 hours incubation at 56°C . The oligos were precipitated by addition of 0.3M Na-Acetate and 2 volumes ethanol. The samples were then centrifuged and the pellets resuspended in either 100 μl or 1ml of water. OD_{260} was determined using a Perkin Elmer Lambda Bio spectrophotometer and the concentration was determined and adjusted to 2-10pmol/ μl .

C) Amplification

The standard PCR protocol was as follows: 50-200ng of genomic DNA were used as a template in the presence of 20-40 μM of each oligo, 400-800 μM dNTPs solution, 1x PCR buffer (including 1.5mM MgCl_2), 2.5 units *TaqI* DNA polymerase (using Perkin-Elmer AmpliTaq, GIBCO Platinum, Pwo DNA polymerase, or Tahara Shuzo Taq polymerase).

In some cases, PCR was optimised by the addition of 10 μl DMSO or 50 μl 2M betaine.

After a hot start (adding the polymerase during a preliminary 3 minute incubation of the whole mix at 95°C), each sample underwent a double-step amplification: the first 5 cycles were performed using as the hybridization temperature the one of the oligos excluding the restriction enzymes tail, followed by 30 cycles performed according to the hybridization temperature of the whole length oligos. The cycles were followed by a final 10 minute extension step at 72°C .

The standard cycles were as follows:

	Denaturation	Hybridisation	Elongation
First 5 cycles	30 seconds	30 seconds	30-60 seconds
	95°C	$50-55^\circ\text{C}$	72°C
Last 30 cycles	30 seconds	30 seconds	30-60 seconds
	95°C	$65-70^\circ\text{C}$	72°C

The elongation time varied according to the length of the ORF to be amplified.

The amplifications were performed using either a 9600 or a 2400 Perkin Elmer GeneAmp PCR System. To check the results, 1/10 of the amplification volume was loaded onto a 1-1.5% agarose gel and the size of each amplified fragment compared with a DNA molecular weight marker.

5 The amplified DNA was either loaded directly on a 1% agarose gel or first precipitated with ethanol and resuspended in a suitable volume to be loaded on a 1% agarose gel. The DNA fragment corresponding to the right size band was then eluted and purified from gel, using the Qiagen Gel Extraction Kit, following the instructions of the manufacturer. The final volume of the DNA fragment was 30µl or 50µl of either water or 10mM Tris, pH 8.5.

D) Digestion of PCR fragments

10 The purified DNA corresponding to the amplified fragment was split into 2 aliquots and double-digested with:

- *NdeI/XhoI* or *NheI/XhoI* for cloning into pET-21b+ and further expression of the protein as a C-terminus His-tag fusion

15 – *BamHI/XhoI* or *EcoRI/XhoI* for cloning into pGEX-KG and further expression of the protein as N-terminus GST fusion.

- *EcoRI/PstI*, *EcoRI/SalI*, *SalI/PstI* for cloning into pGex-His and further expression of the protein as N-terminus His-tag fusion

20 Each purified DNA fragment was incubated (37°C for 3 hours to overnight) with 20 units of each restriction enzyme (New England Biolabs) in either 30 or 40µl final volume in the presence of the appropriate buffer. The digestion product was then purified using the QIAquick PCR purification kit, following the manufacturer's instructions, and eluted in a final volume of 30 or 50µl of either water or 10mM Tris-HCl, pH 8.5. The final DNA concentration was determined by 1% agarose gel electrophoresis in the presence of titrated molecular weight marker.

E) Digestion of the cloning vectors (pET22B, pGEX-KG, pTRC-His A, and pGex-His)

25 10µg plasmid was double-digested with 50 units of each restriction enzyme in 200µl reaction volume in the presence of appropriate buffer by overnight incubation at 37°C. After loading the

whole digestion on a 1% agarose gel, the band corresponding to the digested vector was purified from the gel using the Qiagen QIAquick Gel Extraction Kit and the DNA was eluted in 50µl of 10mM Tris-HCl, pH 8.5. The DNA concentration was evaluated by measuring OD₂₆₀ of the sample, and adjusted to 50µg/µl. 1µl of plasmid was used for each cloning procedure.

- 5 The vector pGEX-His is a modified pGEX-2T vector carrying a region encoding six histidine residues upstream to the thrombin cleavage site and containing the multiple cloning site of the vector pTRC99 (Pharmacia).

F) Cloning

10 The fragments corresponding to each ORF, previously digested and purified, were ligated in both pET22b and pGEX-KG. In a final volume of 20µl, a molar ratio of 3:1 fragment/vector was ligated using 0.5µl of NEB T4 DNA ligase (400 units/µl), in the presence of the buffer supplied by the manufacturer. The reaction was incubated at room temperature for 3 hours. In some experiments, ligation was performed using the Boehringer "Rapid Ligation Kit", following the manufacturer's instructions.

15 In order to introduce the recombinant plasmid in a suitable strain, 100µl *E. coli* DH5 competent cells were incubated with the ligase reaction solution for 40 minutes on ice, then at 37°C for 3 minutes, then, after adding 800µl LB broth, again at 37°C for 20 minutes. The cells were then centrifuged at maximum speed in an Eppendorf microfuge and resuspended in approximately 200µl of the supernatant. The suspension was then plated on LB ampicillin (100mg/ml).

20 The screening of the recombinant clones was performed by growing 5 randomly-chosen colonies overnight at 37°C in either 2ml (pGEX or pTC clones) or 5ml (pET clones) LB broth + 100µg/ml ampicillin. The cells were then pelleted and the DNA extracted using the Qiagen QIAprep Spin Miniprep Kit, following the manufacturer's instructions, to a final volume of 30µl. 5µl of each individual miniprep (approximately 1g) were digested with either *NdeI/XhoI* or *BamHI/XhoI* and the whole digestion loaded onto a 1-1.5% agarose gel (depending on the expected insert size), in parallel with the molecular weight marker (1Kb DNA Ladder, GIBCO). The screening of the
25 positive clones was made on the base of the correct insert size.

G) Expression

Each ORF cloned into the expression vector was transformed into the strain suitable for expression of the recombinant protein product. 1µl of each construct was used to transform 30µl of *E. coli* BL21 (pGEX vector), *E. coli* TOP 10 (pTRC vector) or *E. coli* BL21-DE3 (pET vector), as described above. In the case of the pGEX-His vector, the same *E. coli* strain (W3110) was used for initial cloning and expression. Single recombinant colonies were inoculated into 2ml LB+Amp (100µg/ml), incubated at 37°C overnight, then diluted 1:30 in 20ml of LB+Amp (100µg/ml) in 100ml flasks, making sure that the OD₆₀₀ ranged between 0.1 and 0.15. The flasks were incubated at 30°C into gyratory water bath shakers until OD indicated exponential growth suitable for induction of expression (0.4-0.8 OD for pET and pTRC vectors; 0.8-1 OD for pGEX and pGEX-His vectors). For the pET, pTRC and pGEX-His vectors, the protein expression was induced by addition of 1mM IPTG, whereas in the case of pGEX system the final concentration of IPTG was 0.2mM. After 3 hours incubation at 30°C, the final concentration of the sample was checked by OD. In order to check expression, 1ml of each sample was removed, centrifuged in a microfuge, the pellet resuspended in PBS, and analysed by 12% SDS-PAGE with Coomassie Blue staining. The whole sample was centrifuged at 6000g and the pellet resuspended in PBS for further use.

H) GST-fusion proteins large-scale purification

A single colony was grown overnight at 37°C on LB+Amp agar plate. The bacteria were inoculated into 20ml of LB+Amp liquid culture in a water bath shaker and grown overnight. Bacteria were diluted 1:30 into 600ml of fresh medium and allowed to grow at the optimal temperature (20-37°C) to OD₅₅₀ 0.8-1. Protein expression was induced with 0.2mM IPTG followed by three hours incubation. The culture was centrifuged at 8000rpm at 4°C. The supernatant was discarded and the bacterial pellet was resuspended in 7.5ml cold PBS. The cells were disrupted by sonication on ice for 30 sec at 40W using a Branson sonifier B-15, frozen and thawed twice and centrifuged again. The supernatant was collected and mixed with 150µl Glutathione-Sepharose 4B resin (Pharmacia) (previously washed with PBS) and incubated at room temperature for 30 minutes. The sample was centrifuged at 700g for 5 minutes at 4°C. The resin was washed twice with 10ml cold PBS for 10 minutes, resuspended in 1ml cold PBS, and loaded on a disposable column. The resin was washed twice with 2ml cold PBS until the flow-through reached OD₂₈₀ of 0.02-0.06. The GST-fusion protein was eluted by addition of 700µl cold Glutathione elution buffer (10mM reduced

glutathione, 50mM Tris-HCl) and fractions collected until the OD_{280} was 0.1. 21 μ l of each fraction were loaded on a 12% SDS gel using either Biorad SDS-PAGE Molecular weight standard broad range (M1) (200, 116.25, 97.4, 66.2, 45, 31, 21.5, 14.4, 6.5 kDa) or Amersham Rainbow Marker (M2) (220, 66, 46, 30, 21.5, 14.3 kDa) as standards. As the MW of GST is 26kDa, this value must be added to the MW of each GST-fusion protein.

I) His-fusion solubility analysis

To analyse the solubility of the His-fusion expression products, pellets of 3ml cultures were resuspended in buffer M1 [500 μ l PBS pH 7.2]. 25 μ l lysozyme (10mg/ml) was added and the bacteria were incubated for 15 min at 4°C. The pellets were sonicated for 30 sec at 40W using a Branson sonifier B-15, frozen and thawed twice and then separated again into pellet and supernatant by a centrifugation step. The supernatant was collected and the pellet was resuspended in buffer M2 [8M urea, 0.5M NaCl, 20mM imidazole and 0.1M NaH_2PO_4] and incubated for 3 to 4 hours at 4°C. After centrifugation, the supernatant was collected and the pellet was resuspended in buffer M3 [6M guanidinium-HCl, 0.5M NaCl, 20mM imidazole and 0.1M NaH_2PO_4] overnight at 4°C. The supernatants from all steps were analysed by SDS-PAGE.

J) His-fusion large-scale purification.

A single colony was grown overnight at 37°C on a LB + Amp agar plate. The bacteria were inoculated into 20ml of LB+Amp liquid culture and incubated overnight in a water bath shaker. Bacteria were diluted 1:30 into 600ml fresh medium and allowed to grow at the optimal temperature (20-37°C) to OD_{550} 0.6-0.8. Protein expression was induced by addition of 1mM IPTG and the culture further incubated for three hours. The culture was centrifuged at 8000rpm at 4°C, the supernatant was discarded and the bacterial pellet was resuspended in 7.5ml of either (i) cold buffer A (300mM NaCl, 50mM phosphate buffer, 10mM imidazole, pH 8) for soluble proteins or (ii) buffer B (urea 8M, 10mM Tris-HCl, 100mM phosphate buffer, pH 8.8) for insoluble proteins.

The cells were disrupted by sonication on ice for 30 sec at 40W using a Branson sonifier B-15, frozen and thawed two times and centrifuged again.

For insoluble proteins, the supernatant was stored at -20°C, while the pellets were resuspended in 2ml buffer C (6M guanidine hydrochloride, 100mM phosphate buffer, 10mM Tris-HCl, pH 7.5) and treated in a homogenizer for 10 cycles. The product was centrifuged at 13000rpm for 40 minutes.

Supernatants were collected and mixed with 150µl Ni²⁺-resin (Pharmacia) (previously washed with either buffer A or buffer B, as appropriate) and incubated at room temperature with gentle agitation for 30 minutes. The sample was centrifuged at 700g for 5 minutes at 4°C. The resin was washed twice with 10ml buffer A or B for 10 minutes, resuspended in 1ml buffer A or B and loaded on a disposable column. The resin was washed at either (i) 4°C with 2ml cold buffer A or (ii) room temperature with 2ml buffer B, until the flow-through reached OD₂₈₀ of 0.02-0.06.

- 10 The resin was washed with either (i) 2ml cold 20mM imidazole buffer (300mM NaCl, 50mM phosphate buffer, 20mM imidazole, pH 8) or (ii) buffer D (urea 8M, 10mM Tris-HCl, 100mM phosphate buffer, pH 6.3) until the flow-through reached the O.D₂₈₀ of 0.02-0.06. The His-fusion protein was eluted by addition of 700µl of either (i) cold elution buffer A (300mM NaCl, 50mM phosphate buffer, 250mM imidazole, pH 8) or (ii) elution buffer B (urea 8M, 10mM Tris-HCl, 15 100mM phosphate buffer, pH 4.5) and fractions collected until the O.D₂₈₀ was 0.1. 21µl of each fraction were loaded on a 12% SDS gel.

K) His-fusion proteins renaturation

- 10% glycerol was added to the denatured proteins. The proteins were then diluted to 20µg/ml using dialysis buffer I (10% glycerol, 0.5M arginine, 50mM phosphate buffer, 5mM reduced glutathione, 20 0.5mM oxidised glutathione, 2M urea, pH 8.8) and dialysed against the same buffer at 4°C for 12-14 hours. The protein was further dialysed against dialysis buffer II (10% glycerol, 0.5M arginine, 50mM phosphate buffer, 5mM reduced glutathione, 0.5mM oxidised glutathione, pH 8.8) for 12-14 hours at 4°C. Protein concentration was evaluated using the formula:

$$\text{Protein (mg/ml)} = (1.55 \times \text{OD}_{280}) - (0.76 \times \text{OD}_{260})$$

25 L) His-fusion large-scale purification

500ml of bacterial cultures were induced and the fusion proteins were obtained soluble in buffer M1, M2 or M3 using the procedure described above. The crude extract of the bacteria was loaded

onto a Ni-NTA superflow column (Qiagen) equilibrated with buffer M1, M2 or M3 depending on the solubilization buffer of the fusion proteins. Unbound material was eluted by washing the column with the same buffer. The specific protein was eluted with the corresponding buffer containing 500mM imidazole and dialysed against the corresponding buffer without imidazole.

- 5 After each run the columns were sanitized by washing with at least two column volumes of 0.5 M sodium hydroxide and reequilibrated before the next use.

M) Mice immunisations

- 10 20µg of each purified protein were used to immunise mice intraperitoneally. In the case of ORF 44, CD1 mice were immunised with Al(OH)₃ as adjuvant on days 1, 21 and 42, and immune response was monitored in samples taken on day 56. For ORF 40, CD1 mice were immunised using Freund's adjuvant, rather than Al(OH)₃, and the same immunisation protocol was used, except that the immune response was measured on day 42, rather than 56. Similarly, for ORF 38, CD1 mice were immunised with Freund's adjuvant, but the immune response was measured on day 49.

N) ELISA assay (sera analysis)

- 15 The acapsulated MenB M7 strain was plated on chocolate agar plates and incubated overnight at 37°C. Bacterial colonies were collected from the agar plates using a sterile dracon swab and inoculated into 7ml of Mueller-Hinton Broth (Difco) containing 0.25% Glucose. Bacterial growth was monitored every 30 minutes by following OD₆₂₀. The bacteria were let to grow until the OD reached the value of 0.3-0.4. The culture was centrifuged for 10 minutes at 10000rpm. The supernatant was discarded and bacteria were washed once with PBS, resuspended in PBS
20 containing 0.025% formaldehyde, and incubated for 2 hours at room temperature and then overnight at 4°C with stirring. 100µl bacterial cells were added to each well of a 96 well Greiner plate and incubated overnight at 4°C. The wells were then washed three times with PBT washing buffer (0.1% Tween-20 in PBS). 200µl of saturation buffer (2.7% Polyvinylpyrrolidone 10 in
25 water) was added to each well and the plates incubated for 2 hours at 37°C. Wells were washed three times with PBT. 200µl of diluted sera (Dilution buffer: 1% BSA, 0.1% Tween-20, 0.1% NaN₃ in PBS) were added to each well and the plates incubated for 90 minutes at 37°C. Wells were washed three times with PBT. 100µl of HRP-conjugated rabbit anti-mouse (Dako) serum diluted 1:2000 in dilution buffer were added to each well and the plates were incubated for 90 minutes at

37°C. Wells were washed three times with PBT buffer. 100µl of substrate buffer for HRP (25ml of citrate buffer pH5, 10mg of O-phenildiamine and 10µl of H₂O) were added to each well and the plates were left at room temperature for 20 minutes. 100µl H₂SO₄ was added to each well and OD₄₉₀ was followed. The ELISA was considered positive when OD₄₉₀ was 2.5 times the respective pre-immune sera.

O) FACScan bacteria Binding Assay procedure.

The acapsulated MenB M7 strain was plated on chocolate agar plates and incubated overnight at 37°C. Bacterial colonies were collected from the agar plates using a sterile dracon swab and inoculated into 4 tubes containing 8ml each Mueller-Hinton Broth (Difco) containing 0.25% glucose. Bacterial growth was monitored every 30 minutes by following OD₆₂₀. The bacteria were let to grow until the OD reached the value of 0.35-0.5. The culture was centrifuged for 10 minutes at 4000rpm. The supernatant was discarded and the pellet was resuspended in blocking buffer (1% BSA, 0.4% NaN₃) and centrifuged for 5 minutes at 4000rpm. Cells were resuspended in blocking buffer to reach OD₆₂₀ of 0.07. 100µl bacterial cells were added to each well of a Costar 96 well plate. 100µl of diluted (1:200) sera (in blocking buffer) were added to each well and plates incubated for 2 hours at 4°C. Cells were centrifuged for 5 minutes at 4000rpm, the supernatant aspirated and cells washed by addition of 200µl/well of blocking buffer in each well. 100µl of R-Phicoerytrin conjugated F(ab)₂ goat anti-mouse, diluted 1:100, was added to each well and plates incubated for 1 hour at 4°C. Cells were spun down by centrifugation at 4000rpm for 5 minutes and washed by addition of 200µl/well of blocking buffer. The supernatant was aspirated and cells resuspended in 200µl/well of PBS, 0.25% formaldehyde. Samples were transferred to FACScan tubes and read. The condition for FACScan setting were: FL1 on, FL2 and FL3 off; FSC-H threshold:92; FSC PMT Voltage: E 02; SSC PMT: 474; Amp. Gains 7.1; FL-2 PMT: 539; compensation values: 0.

P) OMV preparations

Bacteria were grown overnight on 5 GC plates, harvested with a loop and resuspended in 10 ml 20mM Tris-HCl. Heat inactivation was performed at 56°C for 30 minutes and the bacteria disrupted by sonication for 10 minutes on ice (50% duty cycle, 50% output). Unbroken cells were removed by centrifugation at 5000g for 10 minutes and the total cell envelope fraction recovered by centrifugation

at 50000g at 4°C for 75 minutes. To extract cytoplasmic membrane proteins from the crude outer membranes, the whole fraction was resuspended in 2% sarkosyl (Sigma) and incubated at room temperature for 20 minutes. The suspension was centrifuged at 10000g for 10 minutes to remove aggregates, and the supernatant further ultracentrifuged at 50000g for 75 minutes to pellet the outer
5 membranes. The outer membranes were resuspended in 10mM Tris-HCl, pH8 and the protein concentration measured by the Bio-Rad Protein assay, using BSA as a standard.

Q) Whole Extracts preparation

Bacteria were grown overnight on a GC plate, harvested with a loop and resuspended in 1ml of 20mM Tris-HCl. Heat inactivation was performed at 56°C for 30 minutes.

10 R) Western blotting

Purified proteins (500ng/lane), outer membrane vesicles (5µg) and total cell extracts (25µg) derived from MenB strain 2996 were loaded on 15% SDS-PAGE and transferred to a nitrocellulose membrane. The transfer was performed for 2 hours at 150mA at 4°C, in transferring buffer (0.3 % Tris base, 1.44 % glycine, 20% methanol). The membrane was saturated by overnight incubation
15 at 4°C in saturation buffer (10% skimmed milk, 0.1% Triton X100 in PBS). The membrane was washed twice with washing buffer (3% skimmed milk, 0.1% Triton X100 in PBS) and incubated for 2 hours at 37°C with mice sera diluted 1:200 in washing buffer. The membrane was washed twice and incubated for 90 minutes with a 1:2000 dilution of horseradish peroxidase labelled anti-mouse Ig. The membrane was washed twice with 0.1% Triton X100 in PBS and developed with
20 the Opti-4CN Substrate Kit (Bio-Rad). The reaction was stopped by adding water.

S) Bactericidal assay

MC58 strain was grown overnight at 37°C on chocolate agar plates. 5-7 colonies were collected and used to inoculate 7ml Mueller-Hinton broth. The suspension was incubated at 37°C on a nutator and let to grow until OD₆₂₀ was 0.5-0.8. The culture was aliquoted into sterile 1.5ml Eppendorf
25 tubes and centrifuged for 20 minutes at maximum speed in a microfuge. The pellet was washed once in Gey's buffer (Gibco) and resuspended in the same buffer to an OD₆₂₀ of 0.5, diluted 1:20000 in Gey's buffer and stored at 25°C.

50µl of Gey's buffer/1% BSA was added to each well of a 96-well tissue culture plate. 25µl of diluted mice sera (1:100 in Gey's buffer/0.2% BSA) were added to each well and the plate incubated at 4°C. 25µl of the previously described bacterial suspension were added to each well. 25µl of either heat-inactivated (56°C waterbath for 30 minutes) or normal baby rabbit complement were added to each well. Immediately after the addition of the baby rabbit complement, 22µl of each sample/well were plated on Mueller-Hinton agar plates (time 0). The 96-well plate was incubated for 1 hour at 37°C with rotation and then 22µl of each sample/well were plated on Mueller-Hinton agar plates (time 1). After overnight incubation the colonies corresponding to time 0 and time 1 hour were counted.

Table II gives a summary of the cloning, expression and purification results.

Example 1

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 1>:

```

1  ..ACACTGTTGT TTGCAACGGT TCAGGCAAGT GCTAACCAAT GAAGAGCAAG
51  AAGAAGATTT ATATTAGAC CCCGTACAAC GCACTGTTGC CGTGTGATA
101 GTCATTCCG ATAAAGAAG CACGGGAGAA AAAGAAAAAG TAGAAGAAAA
151 TTCAGATTGG GCAGTATATT TCAACGAGAA AGGAGTACTA ACAGCCAGAG
201 AAATCACCyT CAAAGCCGGC GACAACCTGA AAATCAAACA AAACGGCACA
251 AACTTCACCT ACTCGCTGAA AAAAGACCTC AcAGATCTGA CCAGTGTGG
301 AACTGAAAAA TTATCGTTTA GCGCAAACGG CAATAAAGTC AACATcACAA
351 GCGACACCAA AGGCTTGAAT TTTGCGAAAG AAACGGCTGG sACGAACGgC
401 GACACCACGG TTCATCTGAA CGGTATTGGT TCGACTTTGA CCGATACGCT
451 GCTGAATACC GGAGCGACCA CAAACGTAAC CAACGACAAAC GTTACCGATG
501 ACGAGAAAAA ACGTGCGGCA AGCGTTAAAG ACGTATTAAA CGTGGCTGG
551 AACATTAAAG GCGTTAAACC CGGTACAACA GCTTCCGATA ACGTTGATT
601 CGTCCGCACT TACGACACAG TCGAGTTCTT GAGCGCAGAT ACGAAAACAA
651 CGACTGTTAA TGTGGAAAGC AAAGACAACG GCAAGAAAAAC CGAAGTTAAA
701 ATCGGTGCGA AGACTTCTGT TATTAAAGAA AAAGAC...
```

This corresponds to the amino acid sequence <SEQ ID 2; ORF40>:

```

1  ..TLLFATVQAS ANQEEQEEDL YLDPVQRTVA VLIVNSDKEG TGEKEKVEEN
51  SDWAYYENK GVLTAAREITX KAGDNLKIKQ NGTNFTYSLK KDLTDLTSVG
101 TEKLSFSANG NKVNITSDTK GLNFAKETAG TNGDITVHLN GIGSTLDTL
151 LNTGATTNVT NDNVTDDEKK RAASVKDVLN AGWNIKGVKP GTTASDNVDF
201 VRTYDTVEFL SADTKTTTVN VESKDNGKKT EVKIGAKTSV IKEKD...
```

Further work revealed the complete DNA sequence <SEQ ID 3>:

```

1  ATGAACAAAA TATACCGCAT CATTGGAAT AGTGCCCTCA ATGCCTGGGT
51  CGTCGTATCC GAGCTCACAC GCAACCACAC CAAACGCGCC TCCGCAACCG
101 TGAAGACCGC CGTATTGGCG ACACTGTTGT TTGCAACGGT TCAGGCAAGT
151 GCTAACAATG AAGAGCAAGA AGAAGATTTA TATTTAGACC CCGTACAACG
201 CACTGTTGCC GTGTTGATAG TCAATTCCGA TAAAGAAGGC ACGGGAGAAA
251 AAGAAAAAGT AGAAGAAAAT TCAGATTGGG CAGTATATT CAACGAGAAA
301 GGAGTACTAA CAGCCAGAGA AATCACCTC AAAGCCGGCG ACAACCTGAA
351 AATCAAACAA AACGGCACAA ACTTCACCTA CTCGCTGAAA AAAGACCTCA
401 CAGATCTGAC CAGTGTGGA ACTGAAAAAT TATCGTTTAG CGCAAACGGC
451 AATAAAGTCA ACATCACAAAG CGACACCAAA GGCTTGAATT TTGCGAAAGA
501 AACGGCTGGG ACGAACGGCG ACACCACGGT TCATCTGAAC GGTATTGGTT
551 CGACTTTGAC CGATACGCTG CTGAATACCG GAGCGACCAC AAACGTAACC
601 AACGACAACG TTACCGATGA CGAGAAAAAA CGTGCGGCAA GCCTTAAAGA
```

5
10
15
20

```

651 CGTATTAAAC GCTGGCTGGA ACATTAAAGG CGTTAAACCC GGTACAACAG
701 CTTCGGATAA CGTTGATTTC GTCCGCACTT ACGACACAGT CGAGTTCTTG
751 AGCGCAGATA CGAAAACAAC GACTGTTAAT GTGGAAAGCA AAGACAACGG
801 CAAGAAAACC GAAGTTAAAA TCGGTGCGAA GACTTCTGTT ATTAAAGAAA
851 AAGACGGTAA GTTGGTTACT GGTAAAGACA AAGGCGAGAA TGGTCTTCT
901 ACAGACGAAG GCGAAGGCTT AGTGACTGCA AAAGAAGTGA TTGATGCAGT
951 AAACAAGGCT GGTGGAGAA TGAAAACAAC AACCGCTAAT GGTCAAACAG
1001 GTCAAGCTGA CAAGTTTGAA ACCGTTACAT CAGGCACAAA TGTAACCTTT
1051 GCTAGTGGTA AAGGTACAAC TGCGACTGTA AGTAAAGATG ATCAAGGCAA
1101 CATCACTGTT ATGTATGATG TAAATGTCGG CGATGCCCTA AACGTCAATC
1151 AGCTGCAAAA CAGCGGTTGG AATTGGGATT CCAAAGCGGT TGCAGGTTCT
1201 TCGGGCAAAG TCATCAGCGG CAATGTTTCG CCGAGCAAGG GAAAGATGGA
1251 TGAAACCGTC AACATTAATG CCGGCAACAA CATCGAGATT ACCCGCAACG
1301 GTAAAAATAT CGACATCGCC ACTTCGATGA CCCCGCAGTT TTCCAGCGTT
1351 TCGCTCGGCG CGGGGGCGGA TGCGCCCACT TTGAGCGTGG ATGGGGACGC
1401 GCTNAGTGTG GGCAGCAAGA AGGACAACAA ACCCGTCCGC ATTACCAATG
1451 TCGCCCCGGG CGTTAAAGAG GGGGATGTTA CAAACGTCGC ACAACTTAAA
1501 GCGTGGGCGC AAAACTTGAA CAACCGCATC GACAATGTGG ACGGCAACGC
1551 GCGTGGGCGC ATCGCCCAAG CGATTGCAAC CGCAGGTCTG GTTCAGGCGT
1601 ATTTGCCCGG CAAGAGTATG ATGGCGATCG GCGGCGGCAC TTATCGCGCG
1651 GAAGCCGGTT ACGCCATCGG CTACTCCAGT ATTTCCGACG GCGGAAATTG
1701 GATTATCAAA GGCACGGCTT CCGGCAATTC GCGGCGCCAT TTCGGTGCTT
1751 CCGCATCTGT CGGTTATCAG TGGTAA

```

This corresponds to the amino acid sequence <SEQ ID 4; ORF40-1>:

25
30
35

```

1 MNKIYRIWN SALNAWVVVS ELTRNHTKRA SATVKTAFLA TLLFATVQAS
51 ANNEEQEEDL YLDPVQRTVA VLIIVNSDEK TGEKEKVEEN SDWAVYFNEK
101 GVLTAREIPL KAGDNLIKIKO NGTNFTYSLK KDLDLTSLVG TEKLSFSANG
151 NKVNITSDTK GLNFAKETAG TNGDITVHLN GIGSTLTDL LNTGATTNVT
201 NDNVTDDEKK RAASVKDVIN AGWNIKGVPK GTTASDNVDF VRTYDTVEFL
251 SADTKTTTVN VESKDNKKKT EVKIGAKTSV IKEKDGKLVK GKDKGENGSS
301 TDEGEGLVTA KEVIDAVNKA GWRMKTITAN GQTQADKFE TVTSGTNVTF
351 ASKGTTTATV SKDDQGNITV MYDQNVGDAL NVNQLQNSGW NLDKAVAGS
401 SGKVISGNVS PSKGMDETV NINAGNNIEI TRNGKNIDIA TSMTPOFSSV
451 SLGAGADAPT LSVGDALNV GSKKDNKPVR ITNVAPGVKE GDVTNVAQLK
501 GQAQNLNRI DNVGDNRAG IAQAIATAGL VQAYLPKSM MAIGGGTYRG
551 EAGYAIGYSS ISDGGNWIIC GTASGNSRGH FGASASVGYQ W*

```

Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 5>:

40
45
50
55
60
65

```

1 ATGAACAAAA TATACCGCAT CATTGGAAT AGTGCCCTCA ATGCCTGNGT
51 CGCCGTATCC GAGCTCACAC GCAACCACAC CAAACGCGCC TCCGCAACCG
101 TGAAAGACGC CGTATTGGCG AACTGTTGT TTGCAACGGT TCAGGCGAAT
151 GCTACCGATG AAGATGAAGA AGAAGAGTTA GAATCCGTAC AACGCTCTGT
201 CGTAGGGAGC ATTCAAGCCA GTATGGAAGG CAGCGGCGAA TTGGAAACGA
251 TATCATTATC AATGACTAAC GACAGCAAGG AATTGTAGA CCCATACATA
301 GTAGTTACCC TCAAAGCCGG CGACAACCTG AAAATCAAAC AAAACACCAA
351 TGAAACACCC AATGCCAGTA GCTTCACCTA CTCGCTGAAA AAAGACCTCA
401 CAGGCCTGAT CAATGTTGAN ACTGAAAAAT TATCGTTTGG CGCAAACGGC
451 AAGAAAGTCA ACATCATAAG CGACACCAA GGCTTGAATT TCGCGAAAGA
501 AACGGCTGGG ACGAACGGCG ACACCACGGT TCATCTGAAC GGTATCGGTT
551 CGACTTTGAC CGATACGCTT GCGGGTTCTT CTGCTTCTCA CGTTGATGCG
601 GGTAAACNAA GTACACATTA CACTCGTGCA GCAAGTATTA AGGATGTGTT
651 GAATGCGGGT TGGAAATATTA AGGSGTTTAA ANNNGGCTCA ACRACGGTTC
701 AATCAGAAAA TGTGATTTTC GTCCGCACTT ACGACACAGT CGAGTTCTTG
751 AGCGCAGATA CGNAAACAAC GACNGTTAAT GTGGAAAGCA AAGACAACGG
801 CAAGAGAACC GAAGTTAAAA TCGGTGCGAA GACTTCTGTT ATTAAAGAAA
851 AAGACGGTAA GTTGGTTACT GGTAAAGGCA AAGGCGAGAA TGGTCTTCT
901 ACAGACGAAG GCGAAGGCTT AGTGACTGCA AAAGAAGTGA TTGATGCAGT
951 AAACAAGGCT GGTGGAGAA TGAAAACAAC AACCGCTAAT GGTCAAACAG
1001 GTCAAGCTGA CAAGTTTGAA ACCGTTACAT CAGGCACAAA TGTAACCTTT
1051 GCTAGTGGTA AAGGTACAAC TGCGACTGTA AGTAAAGATG ATCAAGGCAA
1101 CATCACTGTT ATGTATGATG TAAATGTCGG CGATGCCCTA AACGTCAATC
1151 AGCTGCAAAA CAGCGGTTGG AATTGGGATT CCAAAGCGGT TGCAGGTTCT
1201 TCGGGCAAAG TCATCAGCGG CAATGTTTCG CCGAGCAAGG GAAAGATGGA
1251 TGAAACCGTC AACATTAATG CCGGCAACAA CATCGAGATT AGCCGCAACG
1301 GTAAAAATAT CGACATCGCC ACTTCGATGG CGCCGCAAGT TTCCAGCGTT
1351 TCGCTCGGCG CGGGGGCAGA TGCGCCCACT TTAAGCGTGG ATGACGAGGG
1401 CGGTTGAAT GTCCGACGCA AGGATGCCAA CAAACCCGTC CGCATTACCA

```


1451 ATGTCGCCCC GGGCGTTAAA GANGGGGATG TTACAAACGT CNCACAACTT
 1501 AAAGGCGTGG CGCAAACTT GAACAACCGC ATCGACAATG TGGACGGCAA
 1551 CGCGCGTGCN GGCATCGCCC AAGCGATTGC AACCGCAGGT CTGGTTCAGG
 1601 CGTATCTGCC CGGCAAGAGT ATGATGGCGA TCGGCGGCGG CACTTATCGC
 1651 GGCGAAGCCG GTTACGCCAT CGGCTACTCC AGTATTCCG ACGGCGGAAA
 1701 TTGGATTATC AAAGGCACGG CTTCCGGCAA TTCGCGCGGC CATTTCGGTG
 1751 CTTCCGCATC TGTCCGTTAT CAGTGGTAA

This encodes a protein having amino acid sequence <SEQ ID 6; ORF40a>:

1 MNKIYRIWN SALNAXVAVS ELTRNHTKRA SATVKTAFLA TLLFATVQAN
 51 ATDEDEEEEL ESVQSVVGS IQASMEGSGE LETISLSMTN DSKEFVDPYI
 101 VVTLKAGDNL KIKQNTNENT NASSFTYSLK KDLTGLINVX TEKLSFGANG
 151 KKVNIISDTK GLNFAKETAG TNGDITVHLN GIGSTLDTL AGSSASHVDA
 201 GNKSTHYTRA ASIKDVLNAG WNIKGVKXGS TTGQSENVDF VRTYDTVEFL
 251 SADTXXXXTVN VESKDNGKRT EVKIGAKTSV IKEKDGKLVG GKGKGENGSS
 301 TDEGEGLVTA KEVIDAVNKA GWRMKTITAN GQTGQADKFE TVTSGTNTVF
 351 ASGKGTATV SKDDQGNITV MYDVNVGDAL NVNQLQNSGW NLDKAVAGS
 401 SGKVISGNVS PSKGKMDTV NINAGNNIEI SRNGKNIDIA TSMAPQFSSV
 451 SLGAGADAPT LSVDDGALN VESKDANKPV RITNVAPGVK XGDVTNVXQL
 501 KGVAQNLNRR IDNVGNARA GLAQAIATAG LVQAYLPGKS MMAIGGGTYR
 551 GEAGYAIGYS SISDGGNWII KGTASGNSRG HFGASASVGY QW*

The originally-identified partial strain B sequence (ORF40) shows 65.7% identity over a 254aa overlap with ORF40a:

25 orf40.pep TLLFATVQASANQEEQEEEDLYLDPVQRTVA
 orf40a SALNAXVAVSELTRNHTKRASATVKTAFLATLLFATVQANATDEDEEEEL--ESVQSV-
 30 orf40.pep VLVNSDKEGTGEKEKVEEN--SDWAVYFNEKGVLTAAREITXKAGDNLKIKQ-----GT
 orf40a VGSIQASMEGSGELETISLSMTNDSKEFVDPYIV----VTLKAGDNLKIKQNTNENTNAS
 35 orf40.pep NFTYSLKKDLTDLTSVGTEKLSFANGNKVNITSDTKGLNFAKETAGTNGDITVHLNGIG
 orf40a SFTYSLKKDLTGLINVXTEKLSFGANGKKVNIISDTKGLNFAKETAGTNGDITVHLNGIG
 40 orf40.pep STLTDTLLNTGATTNVDNDVTDDEKKRAASVKDVLNAGWNIGVKPGTTA--SDNVDFV
 orf40a STLTDTLAGSSAS-HVDAGNXST-HYTRAASIKDVLNAGWNIGVKXGTTGQSENVDFV
 45 orf40.pep RTYDTVEFLSADTKTTTVNVESKDNGKTEVKIGAKTSVIKED
 orf40a RTYDTVEFLSADTXXTTTVNVESKDNGKRTVEVKIGAKTSVIKEDGKLVGKGKGENGSSST
 50

The complete strain B sequence (ORF40-1) and ORF40a show 83.7% identity in 601 aa overlap:

55 orf40-1.pep MNKIYRIWNSALNAXVVSELTRNHTKRASATVKTAFLATLLFATVQASANNEEQEEDL
 orf40a MNKIYRIWNSALNAXVAVSELTRNHTKRASATVKTAFLATLLFATVQANATDEDEEEEL
 60 orf40-1.pep YLDPVQRTVAVLVNSDKEGTGEKEKVEEN--SDWAVYFNEKGVLTAAREITLKAGDNLKIK
 : |||: | : : : ||: | : : : : | : : : : |||: |||

```

orf40a      --ESVQRSV-VGSIQASMEGSGELETISLSMTNDSKEFVDPYIV----VTLKAGDNLKIK
              70      80      90      100     110

5  orf40-1.pep 120      130      140      150      160      170
    orf40a      QN-----GTNFTYSLKKDLTDLTSVGTEKLSFSANGNKVNITSDTKGLNFAKETAGTNG
    orf40a      QNTNENTNASSFTYSLKKDLTGLINVXTEKLSFGANGKKVNIISDTKGLNFAKETAGTNG
              120      130      140      150      160      170

10 orf40-1.pep 180      190      200      210      220      230
    orf40a      DTTVHLNGIGSTLTDTLLNTGATTNVNDNDVTDDEKKRAASVKDVLNAGWNIKGVPKPGTT
    orf40a      DTTVHLNGIGSTLTDTLAGSSAS-HVDAGNXST-HYTRAASIKDVLNAGWNIKGVPKXGST
              180      190      200      210      220      230

15 orf40-1.pep 240      250      260      270      280      290
    orf40a      A--SDNVDFVRTYDTEFLSADTKTTTVNVESKDNGKKTEVKIGAKTSVIKEKDGKLVGTG
    orf40a      TGQSENVDVRTYDTEFLSADTKTTTVNVESKDNGKKTEVKIGAKTSVIKEKDGKLVGTG
              240      250      260      270      280      290

20 orf40-1.pep 300      310      320      330      340      350
    orf40a      KDKGENGSSTDEGEGLVTAKEVIDAVNKAGWRMKTTTTANGQTGQADKFETVTSNTVTFFA
    orf40a      KDKGENGSSTDEGEGLVTAKEVIDAVNKAGWRMKTTTTANGQTGQADKFETVTSNTVTFFA
              300      310      320      330      340      350

25 orf40-1.pep 360      370      380      390      400      410
    orf40a      SGKGTATATVSKDDQGNITVMYDVNVGDALNVNQLQNSGWNLDKAVAGSSGKVISGNVSP
    orf40a      SGKGTATATVSKDDQGNITVMYDVNVGDALNVNQLQNSGWNLDKAVAGSSGKVISGNVSP
              360      370      380      390      400      410

30 orf40-1.pep 420      430      440      450      460      470
    orf40a      SKGKMDETVNIAGNNEITRNGKNIDIATSMTPQFSSVSLGAGADAPTLSDVDGD-ALNV
    orf40a      SKGKMDETVNIAGNNEISRNGKNIDIATSMAPQFSSVSLGAGADAPTLSDVDEGALNV
              420      430      440      450      460      470

35 orf40-1.pep 480      490      500      510      520      530
    orf40a      GSKKDNKPVRLTNVAPGVKEGDVTNVAQLKGVAQNLLNNRIDNVGDNARAGIAQAIATAGL
    orf40a      GSKDANKPVRLTNVAPGVKXGVDVTNVXQLKGVAQNLLNNRIDNVGDNARAGIAQAIATAGL
              480      490      500      510      520      530

40 orf40-1.pep 540      550      560      570      580      590
    orf40a      VQAYLPGKSMMAIGGGTYRGEAGYAIGYSSISDGGNWIIGKTASGNSRGHFGASASVGYQ
    orf40a      VQAYLPGKSMMAIGGGTYRGEAGYAIGYSSISDGGNWIIGKTASGNSRGHFGASASVGYQ
              540      550      560      570      580      590

45 orf40-1.pep WX
    orf40a      WX

```

55 Computer analysis of these amino acid sequences gave the following results:

Homology with Hsf protein encoded by the type b surface fibrils locus of *H. influenzae* (accession number U41852)

ORF40 and Hsf protein show 54% aa identity in 251 aa overlap:

```

60 Orf40 1  TLLFATVQASANQEEQEEDLYLDPVQRTVAVLIVNSDXXXXXXXXXXXXXNSDWAIFYNEK 60
    Hsf  41  TLLFATVQA+A E++E LDPV RT VL +SD NS+W +YF+ K
    Orf40 61  GVLTAAREITXKAGDNLKIKQN-----GTNFTYSLKKDLTDLTSVGTEKLSFSANGNKVN 114
    Hsf  96  GVLKAGAITLKAGDNLKIKQNTDESTNASSFTYSLKKDLTDLTSVATEKLSFGANGDKVD 155

```

Orf40 115 ITS DTKGLNFAKETAGTNGDTTVHLNGIGSTLTDTLLNTGAXXXXXXXXXXXXXXKRAAS 174
 ITSD GL AK G+ VHLNG+ STL D + NTG EK RAA+
 Hsf 156 ITSDANGLKLAK-----TGNGNVHLNGLDSTLPDAVTNTGVLSSSFTPNDV-EKTRAAT 209

Orf40 175 VKDVLNAGWNIKGVPKGTASDNVDFVRTYDTVEFLSADTKTTTVNVESKDNGKKTEVKI 234
 VKDVLNAGWNIKG K ++VD V Y+ VEF++ D T V + +K+NGK TEVK
 Hsf 210 VKDVLNAGWNIKGARTAGGNVESVDLV SAYNNVEFITGDKNTLDVVLTAKEKNGKTEVKF 269

Orf40 235 GAKTSVIKEKD 245
 KTSVIKEKD
 Hsf 270 TPKTSVIKEKD 280

ORF40a also shows homology to Hsf:

gi|1666683 (U41852) hsf gene product [Haemophilus influenzae] Length = 2353
 Score = 153 (67.7 bits), Expect = 1.5e-116, Sum P(11) = 1.5e-116
 Identities = 33/36 (91%), Positives = 34/36 (94%)

Query: 16 VAVSELTRNHTKRASATVKTAVLATLLFATVQANAT 51
 V VSELTR HTKRASATV+TAVLATLLFATVQANAT
 Sbjct: 17 VVVSELTRHTTKRASATVETAVLATLLFATVQANAT 52

Score = 161 (71.2 bits), Expect = 1.5e-116, Sum P(11) = 1.5e-116
 Identities = 32/38 (84%), Positives = 36/38 (94%)

Query: 101 VTLKAGDNLKIKONTNENTNASSFTYSLKKDLTG LINV 138
 +TLKAGDNLKIKONT+E+TNASSFTYSLKKDLT L +V
 Sbjct: 103 ITLKAGDNLKIKONTDESTNASSFTYSLKKDLTDLTSV 140

Score = 110 (48.7 bits), Expect = 1.5e-116, Sum P(11) = 1.5e-116
 Identities = 21/29 (72%), Positives = 25/29 (86%)

Query: 138 VTEKLSFGANGKKVNIISDTKGLNFAKET 166
 V++KLS G NG KVNI SDTKGLNFAK++
 Sbjct: 1439 VSDKLSLGTNGKNVITS DTKGLNFAKDS 1467

Score = 85 (37.6 bits), Expect = 1.5e-116, Sum P(11) = 1.5e-116
 Identities = 18/32 (56%), Positives = 20/32 (62%)

Query: 169 TNGDTTVHLNGIGSTLTDTLAGSSSSSHVDAGN 200
 T D +HLNGI STLDTL S A+ GN
 Sbjct: 1469 TGDANIHLNGIASTLTDTLLNSGATTNLGNN 1500

Score = 92 (40.7 bits), Expect = 1.5e-116, Sum P(11) = 1.5e-116
 Identities = 16/19 (84%), Positives = 19/19 (100%)

Query: 206 RAASIKDVLNAGWNIKGVK 224
 RAAS+KDVLNAGWN++GVK
 Sbjct: 1509 RAASVKDVLNAGWNVRGVK 1527

Score = 90 (39.8 bits), Expect = 1.5e-116, Sum P(11) = 1.5e-116
 Identities = 17/28 (60%), Positives = 20/28 (71%)

Query: 226 STTGQSENVDFVRTYDTVEFLSADTTT 253
 S Q EN+DFV TYDTV+F+S D TT
 Sbjct: 1530 SANNQVENIDFVATYDTVDFVSGDKDTT 1557

Based on homology with Hsf, it was predicted that this protein from *N.meningitidis*, and its epitopes, could be useful antigens for vaccines or diagnostics.

ORF40-1 (61kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure

1A shows the results of affinity purification of the His-fusion protein, and Figure 1B shows the

results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for FACS analysis (Figure 1C), a bactericidal assay (Figure 1D), and ELISA (positive result). These experiments confirm that ORF40-1 is a surface-exposed protein, and that it is a useful immunogen.

5 Figure 1E shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF40-1.

Example 2

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 7>

```

1  ATGTTACGTT TGACTGCTTT AGCCGTATGC ACCGCCCTCG CTTTGGGCGC
51  GTGTTGCGCG CAAATTCG ACTCTGCCCC ACAAGCCAAA GaACAGGCGG
101 TTTCCGCCGC ACAAAACGAA GgCGCGTCCG TTACCGTCAA AACCGCGCGC
151 GGGGACGTTT AAATACCGCA AAACCCCGAA CGCATCGCCG TTTACGATT
201 GGGTATGCTC GACACCTTGA GCAAACTGGG CGTGAAAACC GGTTTGTCCG
251 TCGATAAAAA CCGCTGCGG TATTTAGAGG AATATTTCAA AACGACAAAA
301 CCTGCCGGCA CTTTGTTCGA GCCGGATTAC GAAACGCTCA ACGCTTACAA
15  351 ACCGCAGCTC ATCATCATCG GCAGCCGCGC CgCCAAGGCG TTTGACAAAT
401 TGAACGAAAT CGCGCCGACC ATCGTmWTGA CCGCCGATAC CGCCAACCTC
451 AAAGAAAGTG CCAAGAGGCG ATCGACGCTG GCGCAAATCT TC..

```

This corresponds to the amino acid sequence <SEQ ID 8; ORF38>:

```

1  MLRLTALAVC TALALGACSP QNSDSAPQAK EQAVSAAQTE GASVTVKTAR
20  51  GDVQIPQNPE RIAVYDLGML DTLKSLGVKT GLSVDKNRLP YLEEFKTTK
101  PAGTLFEPDY ETLNAYKPQL IIIGSRAAKA FDKLNEIAPT IXXTADTANL
151  KESAKEASTL AQIF..

```

Further work revealed the complete nucleotide sequence <SEQ ID 9>:

```

25  1  ATGTTACGTT TGACTGCTTT AGCCGTATGC ACCGCCCTCG CTTTGGGCGC
51  GTGTTGCGCG CAAATTCG ACTCTGCCCC ACAAGCCAAA GAACAGGCGG
101 TTTCCGCCGC ACAAAACGAA GGCGCGTCCG TTACCGTCAA AACCGCGCGC
151 GGGGACGTTT AAATACCGCA AAACCCCGAA CGCATCGCCG TTTACGATT
201 GGGTATGCTC GACACCTTGA GCAAACTGGG CGTGAAAACC GGTTTGTCCG
251 TCGATAAAAA CCGCTGCGG TATTTAGAGG AATATTTCAA AACGACAAAA
30  301 CCTGCCGGCA CTTTGTTCGA GCCGGATTAC GAAACGCTCA ACGCTTACAA
351 ACCGCAGCTC ATCATCATCG GCAGCCGCGC CGCCAAGGCG TTTGACAAAT
401 TGAACGAAAT CGCGCCGACC ATCGAAATGA CCGCCGATAC CGCCAACCTC
451 AAAGAAAGTG CCAAGAGGCG CATCGACGCG CTGGCGCAAA TCTTCGGCAA
501 ACAGGCGGAA GCCGACAAGC TGAAGGCGGA AATCGACGCG TCTTTTGAAG
35  551 CCGCGAAAAC TGCCGCACAA GGTAAAGGCA AAGGTTTGGT GATTTTGGTC
601 AACGGCGGCA AGATGTCGGC TTTCGGCCCG TCTTCACGCT TGGGCGGCTG
651 GCTGCACAAA GACATCGGCG TTCCCGCTGT CGATGAATCA ATTAAAGAAG
701 GCAGCCACGG TCAGCCTATC AGCTTTGAAT ACCTGAAAGA GAAAAATCCC
751 GACTGGCTGT TTGTCTTGA CCGAAGCGCG GCCATCGGCG AAGAGGGTCA
40  801 GGCGCGGAAA GACGTGTTGG ATAATCCGCT GGTGGCCGAA ACAACCGCTT
851 GGAAAAAGG ACAGGTCGTG TACCTCGTTC CTGAAACTTA TTTGGCAGCC
901 GGTGGCGCGC AAGAGCTGCT GAATGCAAGC AAACAGGTTG CCGACGCTTT
951 TAACGGGCA AAATAA

```

This corresponds to the amino acid sequence <SEQ ID 10; ORF38-1>:

```

45  1  MLRLTALAVC TALALGACSP QNSDSAPQAK EQAVSAAQTE GASVTVKTAR
51  GDVQIPQNPE RIAVYDLGML DTLKSLGVKT GLSVDKNRLP YLEEFKTTK
101  PAGTLFEPDY ETLNAYKPQL IIIGSRAAKA FDKLNEIAPT IEMTADTANL
151  KESAKERIDA LAQIFGKQAE ADKLKAEIDA SFEAAKTAQ GKGKGLVILV
201  NGKMSAFGP SSRLGWLHK DIGVPAVDES IKEGSHGQPI SFEYLKEKNP
50  251 DWLFVLDRSA AIGEEGQAAK DVLDNPLVAE TTAWKKGQVV YLVPETYLAA

```

301 GGAQELLNAS KQVADAFNAA K*

Computer analysis of this amino acid sequence reveals a putative prokaryotic membrane lipoprotein lipid attachment site (underlined).

Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 11>:

```

5      1 ATGTTACGTT TGA CTGCTTT AGCCGTATGC ACCGCCCTCG CTTTGGGCGC
      51 GTGTTGCGCG CAAAATTCCG ACTCTGCCCC ACAAGCCAAA GAACAGGCGG
     101 TTTCCGCGCG ACAATCCGAA GCGGTGTCCG TTACGCTCAA AACGGCGCGC
     151 GCGGATGTTT AAATACCGCA AAACCCCGAA CGTATCGCCG TTTACGATTT
     201 GGGTATGCTC GACACCTTGA GCAAACTGGG CGTGAAAACC GGTTTGTCCG
    10  251 TCGATAAAAA CCGCCTGCCG TATTTAGAGG AATATTTCAA AACGACAAAA
     301 CCTGCCGGAA CTTTGTTCGA GCCGGATTAC GAAACGCTCA ACGCTTACAA
     351 ACCGCAGCTC ATCATCATCG GCAGCCGCGC AGCCAAAGCG TTTGACAAAT
     401 TGAACGAAAT CGCGCCGACC ATCGAAATGA CCGCCGATAC CGCCAACCTC
     451 AAAGAAAGTG CCAAAGAGCG TATCGACGCG CTGGCGCAAA TCTTCGGCAA
    15  501 AAAGCGGGAA GCCGACAAGC TGAAGCGGGA AATCGACGCG TCTTTTGAAG
     551 CCGCGAAAAC TGCCGCGCAA GGCAAGGCA AGGGTTTGGT GATTTTGGTC
     601 AACGCGGGCA AGATGTCCGC CTTCGGCCCG TCTTCACGAC TGGGCGGCTG
     651 GCTGCACAAA GACATCGGCG TTCCCGCTGT TGACGAAGCC ATCAAAGAAG
     701 GCAGCCACGG TCAGCCTATC AGCTTTGAAT ACCTGAAAGA GAAAAATCCC
    20  751 GACTGCGTGT TTGTCTTGA CCGCAGCGCG GCCATCGGCG AAGAGGTCAT
     801 GGCGGCGAAA GACGTGTTGA ACAATCCGCT GGTTCGCGAA ACAACCGCTT
     851 GGAAAAAAGG ACAAGTCGTT TACCTGTGTC CTGAACTTA TTTGGCAGCC
     901 GGTGGCGCGC AAGAGCTACT GAATGCAAGC AAACAGGTTG CCGACGCTTT
     951 TAACGCGGCA AAATAA
  
```

25 This encodes a protein having amino acid sequence <SEQ ID 12; ORF38a>:

```

      1 MLRLTALAVC TALALGACSP QNSDSAPQAK EQAVSAAQSE GVSVTVKTR
     51 GDVQIPQNP ERIAVYDLGML DTL SKLG VKT GLSVDKNR LP YLEEFKTTK
    101 PAGTLFEPDY ETLNAYKPQL IIIGSRAAKA FDKLNEIAPT IEMTADTANL
    151 KESAKERIDA LAQIFGKKA EADKLKAEIDA SFEAAKTAAQ KGKGLVILV
    201 NGGKMSAFGP SSRLGGWLHK DIGVPAVDEA IKEGSHGQPI SFEYLKEKNP
    251 DWLFVLD RSA AIGEEGQAAK DV LNNPLVAE TTAWKKGQV YLVPETLAA
    301 GGAQELLNAS KQVADAFNAA K*
  
```

The originally-identified partial strain B sequence (ORF38) shows 95.2% identity over a 165aa overlap with ORF38a:

```

35      10      20      30      40      50      60
    orf38.pep MLRLTALAVCTALALGACSPQNSDSAPQAKEQAVSAAQTEGASVTVKTRAGDVQIPQNP
    orf38a     MLRLTALAVCTALALGACSPQNSDSAPQAKEQAVSAAQSEGVSVTVKTRAGDVQIPQNP
      10      20      30      40      50      60
    40      70      80      90     100     110     120
    orf38.pep RIAVYDLGMLDTLSKLG VKTGLSVDKNR LPYLEEFKTTK PAGTLFEPDYETLNAYKPQL
    orf38a     RIAVYDLGMLDTLSKLG VKTGLSVDKNR LPYLEEFKTTK PAGTLFEPDYETLNAYKPQL
      70      80      90     100     110     120
    45      130     140     150     160
    orf38.pep IIIGSRAAKAFDKLNEIAPTIXXTADTANLKESAKE-ASTLAQIF
    orf38a     IIIGSRAAKAFDKLNEIAPT IEMTADTANLKESAKERIDALAQIFGKKA EADKLKAEIDA
      130     140     150     160     170     180
    50      190     200     210     220     230     240
    orf38a     SFEAAKTAAQKGKGLVILVNGGKMSAFGPSSRLGGWLHKDIGVPAVDEAIKEGSHGQPI
  
```

55 The complete strain B sequence (ORF38-1) and ORF38a show 98.4% identity in 321 aa overlap:

```

5  orf38a.pep  MLRLTALAVCTALALGACSPQNSDSAPQAKEQAVSAAQSEGVSVTVKTARGDVQIPQNPE
   orf38-1    MLRLTALAVCTALALGACSPQNSDSAPQAKEQAVSAAQTEGASVTVKTARGDVQIPQNPE

10 orf38a.pep  RIAVYDLGMLDTLSKLGVKTGLSVDKNRPLYLEEYFKTTKPAGTLFEPDYETLNAYKPQL
   orf38-1    RIAVYDLGMLDTLSKLGVKTGLSVDKNRPLYLEEYFKTTKPAGTLFEPDYETLNAYKPQL

15 orf38a.pep  IIIGSRAAKAFDKLNEIAPTIEMTADTANLKESAKERIDALAQIFGKAEADKLKAEIDA
   orf38-1    IIIGSRAAKAFDKLNEIAPTIEMTADTANLKESAKERIDALAQIFGKQAEADKLKAEIDA

20 orf38a.pep  SFEAAKTAAGKKGKGLVILVNGGKMSAFGPSSRLGGWLHKDIGVPAVDEAIKEGSHGQPI
   orf38-1    SFEAAKTAAGKKGKGLVILVNGGKMSAFGPSSRLGGWLHKDIGVPAVDESIKEGSHGQPI

   orf38a.pep  SFEYLKEKNPDWLFVLDRAAIGEEGQAADVLNPLVAETTAWKKGQVVYLPETYLAA
   orf38-1    SFEYLKEKNPDWLFVLDRAAIGEEGQAADVLNPLVAETTAWKKGQVVYLPETYLAA

   orf38a.pep  GGAQELLNASKQVADAFNAAK
   orf38-1    GGAQELLNASKQVADAFNAAK

```

Computer analysis of these sequences revealed the following:

25 Homology with a lipoprotein (lipo) of *C. jejuni* (accession number X82427)

ORF38 and lipo show 38% aa identity in 96 aa overlap:

```

30 Orf38: 40  EGASVTVKTARGDVQIPQNPERIAVYDLGMLDTLSKLGVKTGLS-VDKNRPLYLEEYFKT 98
   Lipo: 51  EG S  VK + G+ + P+NP ++ + DLG+LDT  L +  ++ V  LP  + FK
   Orf38: 99  TKPAGTLFEPDYETLNAYKPQLIIIGSRAAKAFDKL 134
   Lipo: 111 KPSVGGVQQVDFEAINALKPDLIISGRQSKFYDKL 146

```

Based on this analysis, it was predicted that this protein from *N. meningitidis*, and its epitopes, could
 35 be useful antigens for vaccines or diagnostics.

ORF38-1 (32kDa) was cloned in pET and pGex vectors and expressed in *E. coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 2A shows the results of affinity purification of the His-fusion protein, and Figure 2B shows the results of expression of the GST-fusion in *E. coli*. Purified His-fusion protein was used to immunise
 40 mice, whose sera were used for Western blot analysis (Figure 2C) and FACS analysis (Figure 2D). These experiments confirm that ORF38-1 is a surface-exposed protein, and that it is a useful immunogen.

Figure 2E shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF38-1.

Example 3

45 The following *N. meningitidis* DNA sequence was identified <SEQ ID 13>:

1 ATGAACTTC TGACCACCGC AATCCTGTCT TCCGCAATCG CGCTCAGCAG
 51 TATGGCTGCC GCCGCTGGCA CGGACAACCC CACTGTTGCA AAAAAAACCG
 101 TCAGCTACGT CTGCCAGCAA GGTAAGTAAC CTACGGCTTC
 151 AACAAACAGG GTCTGACCAC ATACGCTTCC GCCGTCATCA ACGGCAAACG
 201 CGTGCAATG CCTGTCAATT TGGACAAATC CGACAATGTG GAAACATTCT
 251 ACGGCAAAGA AGGCGGTAT GTTTTGGGTA CCGGCGTGAT GGATGGCAAA
 301 TCCTACCGCA AACAGCCAT TATGATTACC GCACCTGACA ACCAAATCGT
 351 CTTCAAAGAC TGTCCCCAC GTTAA

This corresponds to the amino acid sequence <SEQ ID 14; ORF44>:

1 MKLLTTAILS SAIALSSMAA AAGTDNPTVA KKTVSIVCQO GKKVKVTYGF
 51 NKQGLTTYAS AVINGKRVQM PVNLDKSDNV ETFYKKEGGY VLGTGVMDGK
 101 SYRKQIMIT APDNQIVFKD CSPR*

Computer analysis of this amino acid sequence predicted the leader peptide shown underlined.

Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 15>:

1 ATGAACTTC TGACCACCGC AATCCTGTCT TCCGCAATCG CGCTCAGCAG
 51 TATGGCTGCT GCTGCCGGCA CGAACAACCC CACCGTTGCC AAAAAAACCG
 101 TCAGCTACGT CTGCCAGCAA GGTAAGTAAC CTACGGCTTT
 151 AACAAACAGG CCTGTGACCAC ATACGCTTCC GCCGTCATCA ACGGCAAACG
 201 TGTGCAATG CCTGTCAATT TGGACAAATC CGACAATGTG GAAACATTCT
 251 ACGGCAAAGA AGGCGGTAT GTTTTGGGTA CCGGCGTGAT GGATGGCAAA
 301 TCCTATCGCA AACAGCCTAT TATGATTACC GCACCTGACA ACCAAATCGT
 351 CTTCAAAGAC TGTCCCCAC GTTAA

This encodes a protein having amino acid sequence <SEQ ID 16; ORF44a>:

1 MKLLTTAILS SAIALSSMAA AAGTNNPTVA KKTVSIVCQO GKKVKVTYGF
 51 NKQGLTTYAS AVINGKRVQM PVNLDKSDNV ETFYKKEGGY VLGTGVMDGK
 101 SYRKQIMIT APDNQIVFKD CSPR*

The strain B sequence (ORF44) shows 99.2% identity over a 124aa overlap with ORF44a:

		10	20	30	40	50	60
30	orf44.pep	<u>MKLLTTAILSSAIALSSMAA</u> AAGTDNPTVAKKTVSIVCQQGKKVKVTYGFNKQGLTTYAS					
	orf44a	<u>MKLLTTAILSSAIALSSMAA</u> AAGTNNPTVAKKTVSIVCQQGKKVKVTYGFNKQGLTTYAS					
		10	20	30	40	50	60
35	orf44.pep	70	80	90	100	110	120
	orf44a	AVINGKRVQMPVNLDKSDNVETFYKKEGGYVLGTGVMDGKSYRKQIMITAPDNQIVFKD					
		70	80	90	100	110	120
40	orf44.pep	CSPRX					
	orf44a						
		CSPRX					

Computer analysis gave the following results:

Homology with the LecA adhesin of *Eikenella corrodens* (accession number D78153)

ORF44 and LecA protein show 45% aa identity in 91 aa overlap:

Orf44 33 TVSYVCQGGKKVKVTYGFNKQGLTTYASAVINGKRVQMPVNLDKSDNVETFYKKEGGYVL 92
 +V+YVCQGG+++ V Y FN G+ T A +N + +++P NL SDNV+T + GY L
 LecA 135 SVAYVCQGGRRLLNVYRFSAGVPTSaelrvnnrnlrlpynlsasdnvdtvf-SANGYRL 193
 Orf44 93 GTGVMDGKSYRKQIMITAPDNQIVFKDCSP 123
 T MD +YR Q I+++AP+ Q+++KDCSP

LecA 194 TTNAMDSANYRSQDIIVSAPNGQMLYKDCSP 224

Based on homology with the adhesin, it was predicted that this protein from *N.meningitidis*, and its epitopes, could be useful antigens for vaccines or diagnostics.

ORF44-1 (11.2kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 3A shows the results of affinity purification of the His-fusion protein, and Figure 3B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for ELISA, which gave positive results, and for a bactericidal assay (Figure 3C). These experiments confirm that ORF44-1 is a surface-exposed protein, and that it is a useful immunogen.

Figure 3D shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF44-1.

Example 4

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 17>

```

15      1  ..GGCACCGAAT TCAAAACCAC CCTTTCGGGA GCGACATAC AGGCAGGGGT
      51  GGGTGAAAAA GCCCGAGCCG ATGCGAAAT TATCCTAAAA GGCATCGTTA
     101  ACCGCATCCA AACCGAAGAA AAGCTGGAAT CCAACTCGAC CGTATGGCAA
     151  AAGCAGGCCG GAAGCGGCAG CACGGTTGAA ACGCTGAAGC TACCGAGCTT
     201  TGAAGGGCCG GCACTGCCTA AGCTGACCGC TCCCGGCGCG TATATCGCCG
     251  ACATCCCCAA AGGCAACCTC AAAACCGAAA TCGAAAGCT GGCCAAACAG
     20  301  CCCGAATATG CCTATCTGAA ACAGCTTCAG ACGGTCRAGG ACGTGAACCTG
     351  GAACCAAGTA CAGCTCGCTT ACGACAAATG GGACTATAAA CAGGAAGGCC
     401  TAACCGGAGC CGGAGCCGCA ATTANCGCAC TGGCCGTTAC CGTGGTCACC
     451  TCAGGCGCAG GAACCGGAGC CGTATTGGGA TTAANACGNG TGGCCGCCGC
     501  CGCAACCGAT GCAGCATT...
```

25 This corresponds to the amino acid sequence <SEQ ID 18; ORF49>:

```

      1  ..GTEFKTTLSG ADIQAGVGEK ARADAKIILK GIVNRIQTEE KLESNSTVWQ
     51  KQAGSGSTVE TLKLPSFEGP ALPKLTAPGG YIADIPKGNL KTEIEKLAKQ
    101  PEYAYLKQLQ TVKDVNWNQV QLAYDKWDYK QEGLTGAGAA IXALAVTVVT
    151  SGAGTGAVLG LXRVAATAAD AAF..
```

30 Further work revealed the complete nucleotide sequence <SEQ ID 19>:

```

      1  ATGCAACTGC TGGCAGCCGA AGGCATTAC CAACACCAAT TGAATGTTCA
     51  GAAAAGTACC CGTTTCATCG GCATCAAAGT GGGTAAAAGC AATTACAGCA
    101  AAAACGAGCT GAACGAAACC AAAGTCCCGG TACGCGTTAT CGCCCAAACA
    151  GCCAAACCTT GTTCCGGCTG GGATACCGTA CTCGAAGGCA CCGAATTCAA
    201  AACCACCTT TCCGGAGCCG ACATACAGGC AGGGGTGGGT GAAAAAGCCC
    251  GAGCCGATGC GAAAATTATC CTAAAAGGCA TCGTTAACCG CATCCAAACC
    301  GAAGAAAAGC TGGATCCAA CTCGACCGTA TGGCAAAGC AGGCCGGAAG
    351  CGGCAGCAGC GTTGAAACGC TGAAGCTACC GAGCTTTGAA GGGCCGGCAC
    401  TGCCTAAGCT GACCGCTCCC GCGGGCTATA TCGCCGACAT CCCCAAAGGC
    451  AACCTCAAAA CCGAAATCGA AAAGCTGGCC AAACAGCCCG AATATGCCTA
    501  TCTGAAACAG CTTACAGCGG TCAAGGACGT GAACTGGAAC CAAGTACAGC
    551  TCGCTTACGA CAAATGGGAC TATAAACAGG AAGGCCTAAC CGGAGCCGGA
    601  GCCGCAATTA TCGCACTGGC CGTTACCGTG GTCACCTCAG GCGCAGGAAC
    651  CGGAGCCGTA TTGGGATTAA ACGGTGCGGC CGCCGCCGCA ACCGATGCAG
```


701 CATTGCGCTC TTTGGCCAGC CAGGCTTCG TATCGTTCAT CAACAACAAA
 751 GGCAATATCG GTAACACCCCT GAAAGAGCTG GGCAGAAGCA GCACGGTGAA
 801 AAATCTGATG GTTGCCGCTCG CTACCGCAGG CGTAGCCGAC AAAATCGGTG
 851 CTTCCGGCACT GAACAATGTC AGCGATAAGC AGTGGATCAA CAACCTGACC
 901 GTCAACCTGG CCAATGCGGG CAGTGCCGCA CTGATTAATA CCGCTGTCAA
 951 CGGCGGCAGC CTGAAAGACA ATCTGGAAGC GAATATCCTT GCGGCTTTGG
 1001 TGAATACTGC GCATGGAGAG GCAGCAAGTA AAATCAAACA GTTGGATCAG
 1051 CACTACATTG CCCATAAGAT TGCCCATGCC ATAGCGGGCT GTGCGGCAGC
 1101 GCGCGCGAAT AAGGGCAAGT GTCAAGATGG TGCGATCGGT GCGGCGGTCTG
 1151 GTGAAATCCT TGGCGAAACC CTACTGGACG GCAGAGACCC TGGCAGCCTG
 1201 AATGTGAAGG ACAGGGCAAA AATCATTGCT AAGGCGAAGC TGGCAGCAGG
 1251 GCGGTTGCG GCGTTGAGTA AGGGGGATGT GAGTACGGCG GCGAATGCGG
 1301 CTGCTGTGGC GGTAGAGAAT AATTCTTTAA ATGATATACA GGATCGTTTG
 1351 TTGAGTGGAA ATTATGCTTT ATGTATGAGT GCAGGAGGAG CAGAAAGCTT
 1401 TTGTGAGTCT TATCGACCAC TGGGCTTGCC AACTTTGTA AGTGTTCAG
 1451 GAGAAATGAA ATTACCTAAT AAATTCGGGA ATCGTATGTT TAATGGAAAA
 1501 TTAATTATTA AACTAGAAAA TGGCAATGTA TATTTCTCTG TAGGTAAAAAT
 1551 ATGGAGTACT GTAAATCAA CAAATCAAA TATAAGTGGG GTATCTGTCG
 1601 GTTGGGTTTT AAATGTTTCC CCTAATGATT ATTTAAAAGA AGCATCTATG
 1651 AATGATTTCA GAAATAGTAA TCAAAATAAA GCCTATGCAG AAATGATTTT
 1701 CCAGACTTTG GTAGGTGAGA GTGTTGGTGG TAGTCTTTGT CTGACAAGAG
 1751 CCTGCTTTTC GGTAAAGTTCA ACAATATCTA AATCTAAATC TCCTTTTAAA
 1801 GATTCAAAAA TTATTGGGGA AATCGGTTTG GGAAGTGGTG TTGCTGCAGG
 1851 AGTAGAAAAA ACAATATACA TAGGTAACAT AAAAGATATT GATAAATTTA
 1901 TTAGTGCAAA CATAAAAAA TAG

This corresponds to the amino acid sequence <SEQ ID 20; ORF49-1>:

1 MQLLAAEGIH QHQLNVQKST RFIGIKVGKS NYSKNELNET KLPVRVIAQT
 51 AKTRSGWDTV LEGTEFKTTL SGADIQAGVG EKARADAKII LKGIVNRIQT
 101 EEKLESNSTV WQKQAGSGST VETLKLPSFE GPALPKLTAP GGYIADIPKG
 151 NLKTEIEKLA KQPEYAYLKQ LQTVKDVNWN QVQLAYDKWD YKQEGLTGAG
 201 AAIIALAVTV VTSGAGTGAV LGLNGAAAAA TDAAFASLAS QASVSFINNK
 251 GNIGNTLKEL GRSSTVKNLM VAVATAGVAD KIGASALNNV SDKQWINNLT
 301 VNLANAGSAA LINTAVNGGS LKDNLEANIL AALVNTAHGE AASKIKQLDQ
 351 HYIAHKIAHA IAGCAAAAAA KGKQDGAIG AAVGEILGET LLDGRDPGSL
 401 NVKDRAKIIA KAKLAAGAVA ALSKGDVSTA ANAAAVAVEN NSLNDIQDRL
 451 LSGNYALCMS AGGAESFCES YRPLGLPHFV SVSGEMKLPN KFGNRMVNGK
 501 LIINTRNGNV YFSVGKIWSV VKSTKSNISG VSVGWVLNVS PNDYLKEASM
 551 NDFRFSNQNK AYAEMISQTL VGESVGGSLC LTRACFSVSS TISKSKSPFK
 601 DSKIIGEIGL GSGVAAGVEK TIYIGNIKDI DKFISANIKK *

- 40 Computer analysis predicts a transmembrane domain and also indicates that ORF49 has no significant amino acid homology with known proteins. A corresponding ORF from *N.meningitidis* strain A was, however, identified:

ORF49 shows 86.1% identity over a 173aa overlap with an ORF (ORF49a) from strain A of *N.meningitidis*:

45 orf49.pep 10 20 30
 GTEFKTTL SGADIQAGV GEKARADAKI ILK
 orf49a SKNELNETKLPVRVVAQXAATRSWDTV LEGTEFKTTL AGADIQAGVX EKARVD AKI ILK
 50 40 50 60 70 80 90
 orf49.pep GIVNRIQTE EKLESNSTV WQKQAGSGST VETLKLPSFE GPALPKLTAP GGYIADIPKGNL
 orf49a GIVNRIQSE EKLETNSTV WQKQAGSGST IETLKLPSFE SPKLSAPGGYI VDI PKGNL
 55 100 110 120 130 140 150
 orf49.pep KTEIEKLAKQPEYAYLKQLQTVKDVNWNQVQLAYDKWDYKQEGLTGAGAAIXALAVTVVT
 100 110 120 130 140 150

orf49a	KTEIEKLSKQPEYAYLKQLQVAKNINWNQVQLAYDRWDYKQEGLTEAGAAI IALAVTVVT
	160 170 180 190 200 210
5 orf49.pep	SGAGTGAVLGLXRVAATAAF
	:
orf49a	SGAGTGAVLGLNGAXAAATDAAFASLASQASVSFINNKGVDGKTLKELGRSSTVKNLVVA
	220 230 240 250 260 270

ORF49-1 and ORF49a show 83.2% identity in 457 aa overlap:

10 orf49a.pep	XQLLAEGLHKEHLDVQKSRFIGIKVGKSNYSKNEKNETKLPVRVVAQXAATRSQWDTV
orf49-1	MQLLAEGLHQLNVQKSTRFIGIKVGKSNYSKNEKNETKLPVRVIAQTAKTRSGWDTV
15 orf49a.pep	LEGTEFKTTLAGADIQAGVXEKARVDAKIIKLGIVNRIQSEKLETNSTVWQKQAGRGST
orf49-1	LEGTEFKTTLGADIQAGVGEKARADAKIIKLGIVNRIQTEEKLESNSTVWQKQAGSGST
20 orf49a.pep	IETLKLPSFESPTPKLSAPGGYIVDIPKGNLKTEIEKLSKQPEYAYLKQLQVAKNINWN
orf49-1	VETLKLPSFEGPALPKLTAPGGYIADIPKGNLKTEIEKLAKQPEYAYLKQLQTVKDVNWN
25 orf49a.pep	QVQLAYDRWDYKQEGLTEAGAAI IALAVTVVTSGAGTGAVLGLNGAXAAATDAAFASLAS
orf49-1	QVQLAYDKWDYKQEGLTGAGAAI IALAVTVVTSGAGTGAVLGLNGAAAAATDAAFASLAS
30 orf49a.pep	QASVSFINNKGVDGKTLKELGRSSTVKNLVVAATAGVADKIGASALXNVSDKQWNNLT
orf49-1	QASVSFINNKGNIQNTLKELGRSSTVKNLVAVATAGVADKIGASALNNVSDKQWNNLT
35 orf49a.pep	VNLANAGSAALINTAVNGGSLKDXLEANILAALVNTAHGEAASKIKQLDQHYIVHKIAHA
orf49-1	VNLANAGSAALINTAVNGGSLKDNLEANILAALVNTAHGEAASKIKQLDQHYIAHKIAHA
40 orf49a.pep	IAGCAAAAANKGKQDGAIGAAGEIVGEALTNGKNPDTLTAKEREQILAYSKLAVGTVS
orf49-1	IAGCAAAAANKGKQDGAIGAAGEILGETLLDGRDPGSLNVKDRAKI IAKKLAAGAVA
45 orf49a.pep	GVVGGDVNAANAEVAVKNNQLSDXEGREFDNEMTACAKQNPQLCRKNTVKKYQNVAD
orf49-1	ALSKGDVSTAANAAVAVENNSLNDIQDRLLSGNYALCMSAGGAESFCESYRPLGLPHFV
50 orf49a.pep	KRLAASIAICTDISRSTECRTIRKQHLIDSRSLHSSWEAGLIGKDDDEWYKLFKSKSYTOAD
orf49-1	SVSGEMKLPNKFGRNMVNGKLIINTRNGNVYFSVGKIWSTVKSTKSNISGVSVGWVLNVS

45 The complete length ORF49a nucleotide sequence <SEQ ID 21> is:

1	NTGCAACTGC	TGGCAGAAGA	AGGCATCCAC	AAGCACGAGT	TGGATGTCCA
51	AAAAAGCCGC	CGCTTTATCG	GCATCAAGGT	AGGTNAGAGC	AATTACAGTA
101	AAAACGAACT	GAACGAAACC	AAATTGCCTG	TCCGCGTCGT	CGCCCAAANT
151	GCAGCCACCC	GTTCAAGGCTG	GGATACCGTG	CTCGAAGGTA	CCGAATTCAA
201	AACCACGCTG	GCCGGTGCCG	ACATTGAGGC	AGGTGTANGC	GAAAAAGCCC
251	GTGTCCGATG	GAAAATTATC	CTCAAAGGCA	TTGTGAACCG	TATCCAGTCG
301	GAAGAAAAAT	TAGAAACCAA	CTCAACCGTA	TGGCAGAAAC	AGGCCGGACG
351	CGGCAGCACT	ATCGAAACGC	TAAAACTGCC	CAGCTTCGAA	AGCCCTACTC
401	CGCCCAAATT	GTCCGCACCC	GGCGGNTATA	TCGTGACAT	TCCGAAAGGC
451	AAATCTGAAA	CCGAAATCGA	AAAGCTGTCC	AAACAGCCCG	AGTATGCCTA
501	TCTGAACAG	CTCCAAGTAG	CGAAAAACAT	CAACTGGAAT	CAGGTGCAGC
551	TTGCTTACGA	CAGATGGGAC	TACAAACAGG	AGGGCTTAAC	CGAAGCAGGT
601	GCGGCGATTA	TCGCACTGGC	CGTTACCGTG	GTCACCTCAG	GCGCAGGAAC
651	CGGAGCCGTA	TTGGGATTAA	ACGGTGCGNC	CGCCGCCGCA	ACCGATGCAG
701	CATTGCGCTC	TTTGCCGAGC	CAGGCTTCCG	TATCGTTTAT	CAACAACAAA
751	GGCGATGTCTG	GCAAAACCTT	GAAAGAGCTG	GGCAGAAGCA	GCACGGTGAA
801	AAATCTGGTG	GTGCGGCCG	CTACCGCAGG	CGTAGCCGAC	AAAATCGGCG
851	CTTCGGCACT	GANCAATGTC	AGCGATAAGC	AGTGGATCAA	CAACCTGACC
901	GTCAACCTAG	CCAATGCGGG	CAGTGCCGCA	CTGATTATA	CCGCTGTCAA
951	CGGCGGCAGC	CTGAAAGACA	NTCTGGAAGC	GAATATCCTT	GCGGCTTTGG
1001	TCAATACCGC	GCATGGAGAA	GCAGCCAGTA	AAATCAAACA	GTTGGATCAG

1051 CACTACATAG TCCACAAGAT TGCCCATGCC ATAGCGGGCT GTGCGGCAGC
 1101 GCGGCGGAAT AAGGGCAAGT GTCAGGATGG TCGGATAGGT GCGGCTGTGG
 1151 GCGAGATAGT CCGGGAGGCT TTGACAAACG GCAAAATCC TGACACTTTG
 1201 ACAGCTAAAG AACGCGAACA GATTTTGGA TACAGCAAAC TGGTTGCCGG
 5 1251 TACGGTAAGC GGTGTGGTCG GCGGCGATGT AAATGCGGCG GCGAATGCGG
 1301 CTGAGGTAGC GGTGAAAAAT AATCAGCTTA GCGACNAAGA GGGTAGAGAA
 1351 TTTGATAACG AAATGACTGC ATGCGCCAAA CAGAATANTC CTCAACTGTG
 1401 CAGAAAAAAT ACTGTAAAAA AGTATCAAAA TGTTGCTGAT AAAAGACTTG
 1451 CTGCTTCGAT TGCAATATGT ACGGATATAT CCCGTAGTAC TGAATGTAGA
 10 1501 ACAATCAGAA AACACATT TATCGATAGT AGAAGCCTTC ATTCATCTTG
 1551 GGAAGCAGGT CTAATTGGTA AAGATGATGA ATGGTATAAA TTATTCAGCA
 1601 AATCTTACAC CCAAGCAGAT TTGGCTTTAC AGTCTTATCA TTTGAATACT
 1651 GCTGCTAAAT CTTGGCTTCA ATCGGGCAAT ACAAAGCCTT TATCCGAATG
 1701 GATGTCCGAC CAAGGTTATA CACTTATTTC AGGAGTTAAT CCTAGATTCA
 15 1751 TTCCAATACC AAGAGGGTTT GTAAACAAA ATACACCTAT TACTAATGTC
 1801 AATATCCCGG AAGGCATCAG TTTGATACA AACCTANAAA GACATCTGGC
 1851 AAATGCTGAT GGTTTTAGTC AAGAACAGGG CATTAAAGGA GCCCATAACC
 1901 GCACCAATNT TATGGCAGAA CTAAATTCAC GAGGAGGANG NGTAAATCT
 1951 GAAACCCANA CTGATATTGA AGGCATTACC CGAATTAAAT ATGAGATTCC
 20 2001 TACACTAGAC AGGACAGGTA AACCTGATGG TGGATTAAAG GAAATTTCAA
 2051 GTATAAAAC TGTTTATAAT CCTAAAAANT TTTNNGATGA TAAATACTT
 2101 CAAATGGCTC AANATGCTGN TTCACAAGGA TATTCAAAAG CCTCTAAAAT
 2151 TGCTCAAAAT GAAAGAATA AATCAATATC GGAAGAAAAA AATGTCATTC
 2201 AATTCTCAGA AACCTTTGAC GGAATCAAAT TTAGANNNTA TNTNGATGTA
 25 2251 AATACAGGAA GAATTACAAA CATTACCCA GAATAATTTA A

This encodes a protein having amino acid sequence <SEQ ID 22>:

1 XQLLAEEGIH KHELDVQKSR RFIGIKVGXS NYSKNELNET KLPVRVVAQX
 51 AATRSQWDTV LEGTEFKTTL AGADIQAGVX EKARVDKII LKGIVNRIQS
 101 EEKLETNSTV WQKQAGRST IETLKLPSFE SPTPPKLSAP GGYIVDIPKG
 30 151 NLKTEIEKLS KQPEYAYLKQ LQVAKNINWN QVQLAYDRWD YKQEGLTEAG
 201 AAIIALAVTV VTSGAGTGAV LGLNGAXAAA TDAAFASLAS QASVSFINNK
 251 GDVGKTLKEL GRSSTVKNLV VAAATAGVAD KIGASALXNV SDKQWINNLT
 301 VNLNAGSAA LINTAVNGGS LKDXLEANIL AALVNTAHGE AASKIKQLDQ
 351 HYIVHKIAHA IAGCAAAAAN KGKQDGAIG AAVGEIVGEA LTNGKNPDTL
 35 401 TAKEREQILA YSKLVAGTVS GVVGGDVNAA ANAAEVAVKN NQLSDXEGRE
 451 FDNEMTACAK QNXPOLCRKN TVKKYQNVAD KRLAASIAIC TDISRSTECR
 501 TIRKQHLIDS RSLHSSWEAG LIGKDEWYK LFSKSYTQAD LALQSYHLNT
 551 AAKSWLQSGN TKPLSEWMSD QGYTLISGVN PRFIPIPRG VKONTPIITNV
 601 KYPEGISFDT NLXRHLANAD GFSQEQGIKG AHNRTNXMAE LNSRGGXVKS
 40 651 ETXTDIEGIT RIKYEIPTLD RTGKPDGGFK EISSIKTVYN PKFXDCKIL
 701 QMAQXAXSQG YSKASKIAQN ERTKSISERK NVIQFSETFD GIKFRXYXDV
 751 NTGRITNIHP E*

Based on the presence of a putative transmembrane domain, it is predicted that these proteins from *N.meningitidis*, and their epitopes, could be useful antigens for vaccines or diagnostics.

45 Example 5

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 23>

1 ..CGGATCGTTG TAGGTTTGGC GATTTCTTGC GCCGTAGTCA CCGTAGTCCC
 51 AAGTATAACC CAAGGCTTTG TCTTCGCCTT TCATTCCGAT AAGGGATATG
 101 ACGCTTTGGT CCGTATAGCC GTCTTGGGAA CCTTTGTCCA CCCAACGCAT
 50 151 ATCTGCCTGC GGATTCTCAT TGCCGCTTCT TGGCTGCTGA TTTTCTGCC
 201 TTCGCGTTTT TCAACTTCGC GCTTGAGGGC TTCGGCATAT TTGTCGGCCA
 251 ACGCCATTTC TTTCGGATGC AGCTGCCTAT TGTTCCAATC TACATTGCGA
 301 CCCACCACAG CACCACCACT ACCACCAGTT GCATAG

This corresponds to the amino acid sequence <SEQ ID 24; ORF50>:

55 1 ..RIVVGLRISC AVTVVPSIT QGFVFAFHS D KGYDALVGIA VLGTFVHPTH
 51 ICLRILIAAS WLLIFLPSRF STSRLRASAY LSANAISFGC SCLLFQSTFA
 101 PTTAPPLFPV A*

Computer analysis predicts two transmembrane domains and also indicates that ORF50 has no significant amino acid homology with known proteins.

Based on the presence of a putative transmembrane domain, it is predicted that this protein from *N.meningitidis*, and its epitopes, could be useful antigens for vaccines or diagnostics.

5 Example 6

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 25>

```

1  ..AAGTTTGACT TTACCTGGTT TATTCCGGCG GTAATCAAAT ACCGCCGGTT
51  GTTTTTTGAA GTATTGGTGG TGTCCGGTGT GTTGCAGCTG TTTGCGCTGA
101 TTACGCCTCT GTTTTCCCAA GTGGTGATGG ACAAGGTGCT GGTACATCGG
151 GGATTCTCTA CTTTGGATGT GGTGTCGGTG GCCTTGTTGG TGGTGTGCGT
201 GTTTGAGATT GTGTTGGGCG GTTTGCGGAC GTATCTGTTT GCACATACGA
251 CTTACCGTAT TGATGTGGAA TTGGGCGCGC GTTTGTTCCG GCATCTGCTT
301 TCCCTGCCTT TATCTATTT CGAGCACAGA CGAGTGGGTG ATACGGTGGC
351 TCGGGTGCGG GAATTGGAGC AGATTGCGAA TTTCTTGACC GGTGAGGCGC
401 TGACTTCGGT GTTGGATTG GCGTTTTCGT TTATCTTCTT GCGGTGATG
451 TGGTATTACA GCTCCACTCT GACTTGGGTG GTATTGGCTT CGTTG.....
//
1451 ..... ..ATTGCGC
1501 ..... ..ATTGCGC
20 1551 CAACCGGACG GTGCTGATTA TCGCCACCG TCTGTCCACT GTTAAACGG
1601 CACACCGGAT CATTGCCATG GATAAAGGCA GGATTGTGGA AGCGGGAACA
1651 CAGCAGGAAT TGCTGGCGAA CG..AACGGA TATTACCGCT ATCTGTATGA
1701 TTTACAGAAC GGGTAG

```

This corresponds to the amino acid sequence <SEQ ID 26; ORF39>:

```

25 1  ..KFDFTFWIPA VIKYRRLFFE VLVVSVVLQL FALITPLFFQ VMDKVLVHR
51  GFSTLVVSV ALLVVSLEFI VLGLRLTYLF AHTTSRIDVE LGARLFRHLL
101 SLPLSYFEHR RVGDTVARVR ELEQIRNFLT QOALTSVLDL AFSFIFLAVM
151 WYYSSTLTWV VLASL.....
//
30 501 ..... ICAVRT VLIIAHRLST VKTAHRIAM DKGRIVEAGT
551 QQELLANXNG YYRYLYDLQN G*

```

Further work revealed the complete nucleotide sequence <SEQ ID 27>:

```

35 1  ATGTCTATCG TATCCGCACC GCTCCCGGCC CTTCCCGCCC TCATCATCCT
51  CGCCCATTAC CACGGCATTG CCGCCAATCC TGCCGATATA CAGCATGAAT
101 TTTGTACTTC CGCACAGAGC GATTTAAATG AAACGCAATG GCTGTTAGCC
151 GCCAAATCTT TGGGATTGAA GGCAAAGGTA GTCCGCCAGC CTATTAAACG
201 TTTGGCTATG GCGACTTTAC CCGCATTGGT ATGGTGTGAT GACGGCAACC
251 ATTTTCATTTT GGCCAAAACA GACGGTGAGG GTGAGCATGC CCAATTTTGT
40 301 ATACAGGATT TGGTTACGAA TAAGTCTGCG GTATTGTCTT TTGCCGAATT
351 TTCTAACAGA TATTCCGGCA AACTGATATT GGTGCTTCC CGCGCTTCGG
401 TATTGGGCAG TTTGGCAAAG TTTGACTTTA CCTGGTTTAT TCCGGCGGTA
451 ATCAAATACC GCCGGTTGTT TTTGAAGTA TTGGTGGTGT CGTGTTGTTT
501 GCAGCTGTTT CGCTGATTA CGCCTCTGTT TTTCCAAGTG GTGATGGACA
551 AGGTGCTGGT ACATCGGGGA TTCTCTACTT TGGATGTGGT GTCGGTGGCT
45 601 TTGTTGGTGG TGTGCTGTT TGAGATTGTG TTGGGCGGTT TCGGACGTA
651 TCTGTTTGCA CATACCACTT CACGTATTGA TGTGGAATTG GCGCGCGGTT
701 TGTTCGGGCA CTCGCTTCC CTGCTTTTAT CCTATTTTCA GCACAGACGA
751 GTGGGTGATA CGGTGGCTCG GGTGCGGGAA TTGGAGCAGA TTCGCAATT
801 CTTGACCGGT CAGGCGCTGA CTTGGTGTG GGATTGGCG TTTTCGTTTA
50 851 TCTTTCTGGC GGTGATGTGG TATTACAGCT CCACTCTGAC TTGGGTGGTA
901 TTGGCTTCGT TGCCTGCCTA TCGGTTTTGG TCGGCATTTA TCAGTCCGAT
951 ACTGCGGACG CGTCTGAACG ATAAGTTCGC GCGCAATGCA GACAACCACT
1001 CGTTTTTAGT AGAAAGCATC ACTGCGGTGG GTACGGTAAA GCGCATGGCG
1051 GTGGAGCCGC AGATGACGCA GCGTTGGGAC AATCAGTTGG CGGCTTATGT

```

5
10
15
20

```

1101 GGCTTCGGGA TTTCGGGTAA CGAAGTTGGC GGTGGTCGGC CAGCAGGGGG
1151 TGCAGCTGAT TCAGAAGCTG GTGACGGTGG CGACGTTGTG GATTGGCGCA
1201 CGGCTGGTAA TTGAGAGCAA GCTGACGGTG GGGCAGCTGA TTGCGTTTAA
1251 TATGCTCTCG GGACAGGTGG CGGCGCCTGT TATCCGTTTG CGGCAGTTGT
1301 GGCAGGATTT CCAGCAGGTG GGGATTTCGG TGGCGCGTTT GGGGGATATT
1351 CTGAATGCCG CGACCGAGAA TGGCTCTTCG CATTGGGCTT TGCCCGATAT
1401 CCGGGGGGAG ATTACGTTTC AACATGTCTG TTTCCGCTAT AAGCGGGACG
1451 GCAGGCTGAT TTTGCAGGAT TTGAACCTGC GGATTCGGGC GGGGGAAGTG
1501 CTGGGGATTG TGGGACGTTT GGGGTTCGGC AAATCCACAC TCACCAAATT
1551 GGTGCAGCGT CTGTATGTAC CGGAGCAGGG ACGGGTGTG GTGGACGGCA
1601 ACGATTTGGC TTTGGCCGCT CCTGCCTGGC TGGCGCGGCA GGTGCGCGTG
1651 GTCTTGACAG AGAATGTGCT GCTCAACCGC AGCATACCGC ACAATATCGC
1701 GCTGACGGAT ACGGGTATGC CGCTGGAACG CATATCGGAA CGACCCAAAC
1751 TGGCGGGGCG ACACGAGTTT ATTATGGAGC TGCCGGAAGG CTACGGCACC
1801 GTGGTGGGCG AACAAAGGGC CGGCTTGTTC GCGGACAGC GGCAGCGTAT
1851 TGCGATTGCC CGCGCGTTAA TCACCAATCC GCGCATTCTG ATTTTGTATG
1901 AAGCCACGAG CGCGCTGGAT TATGAAAGTG AACGAGCGAT TATCAGAAC
1951 ATGCAGGCCA TTTGCGCCAA CCGGACGGTG CTGATTATCG CCCACCGTCT
2001 GTCCACTGTT AAAACGGCAC ACCGGATCAT TGCCATGGAT AAAGGCAGGA
2051 TTGTGGAAGC GGAACACACG CAGGAATTGC TGGCGAAGCC GAACGGATAT
2101 TACCGCTATC TGTATGATT ACAGAACGGG TAG

```

This corresponds to the amino acid sequence <SEQ ID 28; ORF39-1>:

25
30
35

```

1 MSIVSAPLPA LSALIILAHY HGIAANPADI QHEFCTSAQS DLNETQWLLA
51 AKSLGLKARV VRQPIKRLAM ATLPALVWCD DGNHFILAKT DGEGEHAQFL
101 IQDLVTNKA VLSFAEFSNR YSGKLILVAS RASVLGSLAK FDFTFWIPAV
151 IKYRRLEFEV LVVSVVLQLE ALITPLFFQV VMDKVLVHRG FSTLDVVSVA
201 LLVVSLFEIV LGLLRTYLFA HTTSRIDVEL GARLFRHLLS LPLSYFEHRR
251 VGDTVARVRE LEQIRNFLTQ QALTSVLDLA FSFIFLAVMW YYSSTLTWV
301 LASLPAYAFW SAFISPIRLT RLNDKFARNA DNQSFIVESI TAVGTVKAMA
351 VEPQMTQRWD NQLAAYVASG FRVTKLAVVG QQGVQLIQKL VTVATLWIGA
401 RLVIKSLTV GQLIAFNMLS GQVAAPVIRL AOLWQDFQOV GISVARLGDI
451 LNAPTENASS HLALPDIRGE ITFEHVDFRY KADGRLIQD LNLIRIRAGEV
501 LGIVGRSGSG KSTLTCLVQR LYVPEQGRVL VDGNDLALAA PAWLRRQGV
551 VLQENVLLNR SIRDNIATD TGMPLERIE AAKLAGAHEF IMELPEGYGT
601 VVGEQAGLS GGQRQRIATA RALITNPRIL IFDEATSALD YESERAIMQN
651 MQAICANRTV LIIAHLSTV KTAHRIIAMD KGRIVEAGTQ QELLAKPNGY
701 YRYLYDLQNG *

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

40 ORF39 shows 100% identity over a 165aa overlap with an ORF (ORF39a) from strain A of *N. meningitidis*:

45
50
55
60

```

          10      20      30
orf39.pep      KFDFTWFIPIAVIKYRRLEFEVLVVSVLQQL
          |||||
orf39a      AVLSFAEFSNRYSGKLILVASRASVLGSLAKFDFTWFIPIAVIKYRRLEFEVLVVSVLQQL
          110      120      130      140      150      160

          40      50      60      70      80      90
orf39.pep      FALITPLFFQVVMKDKVLVHRGFSTLDVVSVALLVVSLFEIVLGGRLTYLFAHTTSRIDVE
          |||||
orf39a      FALITPLFFQVVMKDKVLVHRGFSTLDVVSVALLVVSLFEIVLGGRLTYLFAHTTSRIDVE
          170      180      190      200      210      220

          100      110      120      130      140      150
orf39.pep      LGARLFRHLLSLPLSYFEHRRVGDTVARVRELEQIRNFLTQALTSVLDLAFSFIPLAVM
          |||||
orf39a      LGARLFRHLLSLPLSYFEHRRVGDTVARVRELEQIRNFLTQALTSVLDLAFSFIPLAVM
          230      240      250      260      270      280

          160      170      180      190      200      210
orf39.pep      WYYSSTLTWVVLASLXXXXXXXXXXXXXXXXXXXXXXXXXXXXXICANRTVLIIAHLSTV

```

orf39a
 WYSSSTLTWVVLASLPAYAFWSAFISPILRTRLNDKFARNADNQSFLVESITAVGTVKAM
 290 300 310 320 330 340

ORF39-1 and ORF39a show 99.4% identity in 710 aa overlap:

5	orf39-1.pep	MSIVSAPLPALSALIILAHYHGIAANPADIQHEFCTSAQSDLNETQWLLAAKSLGLKAKV
	orf39a	MSIVSAPLPALSALIILAHYHGIAANPADIQHEFCTSAQSDLNETQWLLAAKSLGLKAKV
10	orf39-1.pep	VRQPIKRLAMATLPALVWCDDGNHFILAKTDGEGEHAQFLIQDLVTNKSAVLSFAEFSNR
	orf39a	VRQPIKRLAMATLPALVWCDDGNHFILAKTDGEGEHAQFLIQDLVTNKSAVLSFAEFSNR
15	orf39-1.pep	YSGKLILVASRASVLSLAKFDFTWFI PAVIKYRRLFFEVLVSVVLQLEFALITPLFFQV
	orf39a	YSGKLILVASRASVLSLAKFDFTWFI PAVIKYRRLFFEVLVSVVLQLEFALITPLFFQV
20	orf39-1.pep	VMDKVLVHRGFSTLDVVSVALLVVSLFEIVLGGRLTYLFAHTTSRIDVELGARLFRHLLS
	orf39a	VMDKVLVHRGFSTLDVVSVALLVVSLFEIVLGGRLTYLFAHTTSRIDVELGARLFRHLLS
25	orf39-1.pep	LPLSYFEHRRVGDTVARVRELEQIRNFLTGOALTSVLDLAFSFI FLAVMWYSSSTLTWVV
	orf39a	LPLSYFEHRRVGDTVARVRELEQIRNFLTGOALTSVLDLAFSFI FLAVMWYSSSTLTWVV
30	orf39-1.pep	NQLAAYVASGFRVTKLAVVGQGVQLIQKLVTVATLWIGARLVIESKLTVGQLIAFNMLS
	orf39a	NQLAAYVASGFRVTKLAVVGQGVQLIQKLVTVATLWIGARLVIESKLTVGQLIAFNMLS
35	orf39-1.pep	GQVAAPVIRLAQLWQDFQQVGISVARLGDILNAPTENASSHLALPDIRGEITFEHVDFRY
	orf39a	GQVAAPVIRLAQLWQDFQQVGISVARLGDILNAPTENASSHLALPDIRGEITFEHVDFRY
40	orf39-1.pep	KADGRILQLDLNLRIRAGEVLGIVGRSGSGKSTLTKLVQRLYVPEQGRVLVDGNDLALAA
	orf39a	KADGRILQLDLNLRIRAGEVLGIVGRSGSGKSTLTKLVQRLYVPAQGRVLVDGNDLALAA
45	orf39-1.pep	PAWLRRQGVVVLQENVLNLNRSIRDNIALTDTGMPLERIEAAKLAGAHEFIMELPEGYGT
	orf39a	PAWLRRQGVVVLQENVLNLNRSIRDNIALTDTGMPLERIEAAKLAGAHEFIMELPEGYGT
50	orf39-1.pep	VVGEQGAGLSGGQRORIAARALITNPRIIFDEATSALDYESERAIMQNMQAICANRTV
	orf39a	VVGEQGAGLSGGQRORIAARALITNPRIIFDEATSALDYESERAIMQNMQAICANRTV
	orf39-1.pep	LIIAHRLSTVKTAHRIIAMDKGRIVEAGTQOELLAKPNGYYRYLYDLQNGX
	orf39a	LIIAHRLSTVKTAHRIIAMDKGRIVEAGTQOELLAKPNGYYRYLYDLQNGX

The complete length ORF39a nucleotide sequence <SEQ ID 29> is:

1	ATGTCTATCG	TATCCGCACC	GCTCCCCGCC	CTTCCGCCC	TCATCATCCT
55	51	CGCCCAATAC	CACGGCATTG	CGCCCAATCC	TGCCGATATA
	101	TTGTACTTC	CGCACAGAGC	GATTAAATG	AAACGCAATG
	151	GCCAAATCTT	TGGGATTGAA	GGCAAAGTA	GTCGCCAGC
	201	TTGGCTATG	GCGACTTTAC	CCGCATTGGT	ATGGTGTGAT
	251	ATTTTATTTT	GGCTAAAACA	GACGGTGCGG	GTGAGCATGC
	301	ATACAGGATT	TAACACGAA	TAAGTCTGCG	GTATTGTCTT
60	351	TTCTAACAGA	TATTCGGGCA	AACTGATATT	GGTGTCTTCC
	401	TATGGGCAG	TTTGGCAAAG	TTTGACTTTA	CCTGGTTTAT
	451	ATCAAATACC	GCCGGTTGTT	TTTGAAGTA	TTGGTGGTGT
	501	GCAGCTGTTT	GCGCTGATTA	CGCCTCTGTT	TTTCCAAGTG
	551	AGGTGCTGGT	ACATCGGGGA	TTCTCTACTT	TGGATGTGGT
65	601	TTGTTGGTGG	TGTCGCTGTT	TGAGATTGTG	TTGGGCGGTT
	651	TCTGTTTGCA	CATACGACTT	CACGTATTGA	TGTGGAATTG

701 TGTTCGGCA TCTGCTTCC CTGCTTTAT CCTATTTGCA GCACAGACGA
 751 GTGGGTGATA CGGTGGCTCG GGTGCGGGAA TTGGAGCAGA TTCGCAATTT
 801 CTTGACCGGT CAGGCGCTGA CTTGCGGTGT GGATTGCGCG TTTCGTTTA
 851 TCTTCTGGC GGTGATGTGG TATTACAGCT CCACTCTGAC TTGGGTGGTA
 901 TTGGCTTCGT TGCCTGCCTA TCGCTTTTGG TCGGCATTTA TCAGTCCGAT
 951 ACTGCGGACG CGTCTGAACG ATAAGTTCCG GCGCAATGCA GACAACCAGT
 1001 CGTTTTTAGT AGAAAGCATC ACTGCGGTGG GTACGGTAAA GCGATGGCG
 1051 GTGGAGCCGC AGATGACGCA GCGTTGGGAC AATCAGTTGG CGGCTTATGT
 1101 GGCTTCGGGA TTTCGGGTAA CGAAGTTGGC GGTGGTCGGC CAGCAGGGGG
 1151 TGCAGCTGAT TCAGAAGCTG GTGACGGTGG CGACGTTGTG GATTGGCGCA
 1201 CGGCTGGTAA TTGAGAGCAA GCTGACGGTG GGGCAGCTGA TTGCGTTTAA
 1251 TATGCTCTCG GGACAGGTGG CCGCGCCTGT TATCCGTTT GCGCAGTTGT
 1301 GGCAGGATTT CCAGCAGGTG GGGATTTCCG TGGCGCGTTT GGGGATATT
 1351 CTGAATGCGC CGACCGAGAA TCGCTCTTCG CATTTGGCTT TGCCCGATAT
 1401 CCGGGGGGAG ATTACGTTTC AACATGTGCA TTCCCGCTAT AAGGCGGACG
 1451 GCAGGCTGAT TTGACAGGAT TTGAACCTGC GGATTCGGGC GGGGGAAGTG
 1501 CTGGGATTTG TGGGACGTTT GGGGTCGGC AAATCCACAC TCACCAAAAT
 1551 GGTGCAGCGT CTGTATGTAC CCGCGCAGGG ACGGGTGTG GTGACGGCA
 1601 ACGATTTGGC TTGGCCGCT CCGCTTGGC TCGGCGGCA GGTGCGCGTG
 1651 GTCTTGACG AGAATGTGCT GCTCAACCGC AGCATACGCG ACAATATCGC
 1701 GCTGACGGAT ACGGGTATGC CGCTGGAACG CATTATCGAA GCAGCCAAAC
 1751 TGGCGGGGCG ACACGAGTTT ATTATGGAGC TGCCGGAAGG CTACGGCACC
 1801 GTGGTGGGCG AACAAGGGCG CCGCTTGTG GCGGACAGC GGCAGCGTAT
 1851 TCGGATTGCG CGCGCGTTAA TCACCAATCC GCGCATTCTG ATTTTGTATG
 1901 AAGCCACCAG CGCGCTGGAT TATGAAAGTG AACGAGCGAT TATGCAGAAC
 1951 ATCGAGGCCA TTTCGCCCAA CCGGACGGTG CTGATTATCG CCCACCGTCT
 2001 GTCCACTGTT AAAACGGCAC ACCGGATCAT TGCCATGGAT AAAGGCAGGA
 2051 TTGTGGAAGC GGGAAACACAG CAGGAATTGC TGGCGAAGCC GAACGGATAT
 2101 TACCGCTATC TGTATGATTT ACAGAACGGG TAG

30 This encodes a protein having amino acid sequence <SEQ ID 30>:

1 MSIVSAPLPA LSALIILAHY HGIAANPADI QHEFCTSAQS DLNETQWLLA
 51 AKSLGLKAKV VRQPIKRLAM ATLPALVWCD DGNHFILART DGGGEHAQYL
 101 IQDLTTNKSA VLSFAEFSNR YSGKLILVAS RASVLGSLAK FDFWFIPIAV
 151 IKYRRLFEEV LVSVVLQLE ALITPLFFQV VMDKVLVHRG FSTLDVVSVA
 201 LLVSLFEIV LGLRITYLFA HTTSRIDVEL GARLFRHLLS LPLSYFEHRR
 251 VGDTVARVRE LEQIRNFTLG QALTSVLDLA FSFIFLAVMW YYSSTLTWVV
 301 LASLPAYAFW SAFISPIIRT RLNDKFARNA DNQSFIVESI TAVGTVKAMA
 351 VEPQMTQRWD NQLAAYVASG FRVTKLAVVG QQGVQLIQKL VTATLWIGA
 401 RLVIKSLTV GQLIAFNMLS GQVAAPVIRL AQLWQDFQOV GISVARLGDI
 451 LNAPTENASS HLALPDIRGE ITFEHVDFRY KADGRILIQD LNLIRAGEV
 501 LGIVGRSGSG KSTLTCLKVQR LYVPAQGRVL VDGNLALAA PAWLRRQGVV
 551 VLQENVLLNR SIRDNIATD TGMPLERIE AAKLAGAHEF IMELPEGYGT
 601 VLEQGAGLS GGQRQIAIA RALITNPRIL IFDEATSALD YESERAIMQN
 651 MQAICANRTV LIIAHLSTV KTAHRIIAMD KGRIVEAGTQ QELLAKPNGY
 701 YRYLYDLQNG *

ORF39a is homologous to a cytolysin from *A. pleuropneumoniae*:

sp|P26760|RT1B ACTPL RTX-I TOXIN DETERMINANT B (TOXIN RTX-I SECRETION ATP-
 BINDING PROTEIN) (APX-IB) (HLY-IB) (CYTOLYSIN IB) (CLY-IB)
 >gi|97137|pir||D43599 cytolysin IB - Actinobacillus pleuropneumoniae (serotype 9)
 >gi|38944 (X61112) ClyI-B protein [Actinobacillus pleuropneumoniae] Length = 707
 Score = 931 bits (2379), Expect = 0.0
 Identities = 472/690 (68%), Positives = 540/690 (77%), Gaps = 3/690 (0%)

Query: 20 YHGIAANPADIQHEFCTSAQSDLNETQWXXXXXXXXXXVVRQPIKRLAMATLPALVWC 79
 YH IA NP +++H+F + L+ T W V++ I RLA LPALVW
 Sbjct: 20 YHNIHAVNPEELKHKFDLEGKG-LDLTAWLLAAKSLELAKQVKKAIIDRLAFIALPALVWR 78

Query: 80 DDGNHFILAKTDGGGEHAQYLIQDLTTNKSAVLSFAEFSNRYSGLKILVASRASVLGSLA 139
 +DG HFIL K D E +YLI DL T+ +L AEF + Y GKLILVASRAS++G LA
 Sbjct: 79 EDGKHFILTKIDN--EAKKYLI FDLTHNPRILEQAEFESLYQGKLILVASRASIVGKLA 136

Query: 140 KFDFTFWFIPIAVIKYRXXXXXXXXXXXXXXXXXITPLFFQVVMKVLVHRGFXXXXXXXXX 199
 KFDFTFWFIPIAVIKYR+ ITPLFFQVVMKVLVHRGF
 Sbjct: 137 KFDFTFWFIPIAVIKYRKIFETLIVSIFLQIFALITPLFFQVVMKVLVHRGFSTLNVITV 196

Query: 200 XXXXXXFEIVLGLRITYLFAHTTSRIDVELGARLFRHLLSLPLSYFEHRRVGDTVARVR 259

FEIVL GLRTY+FAH+TSRIDVELGARLFRHLL+LP+SYFE+RRVGD TVARVR
 Sbjct: 197 ALAIVVLF EIVLNGLR TYIFAHSTSRIDVELGARLFRHLLALPISYFENRRVGD TVARVR 256

Query: 260 ELEQIRNFLT GQALTSVLDLAFSFI FLAVMWYSSSTLTWVVLASLPAYAFWSAFISPI LR 319
 EL+QIRNFLT GQALTSVLDL FSFIF AVMWYYS LT V+L SLP Y WS FISPI LR
 Sbjct: 257 ELDQIRNFLT GQALTSVLDLMSFSFI FFAVMWYSPKLT LVILGSLP FFMGWSIFISPI LR 316

Query: 320 TRLNDFARNADNQSF LVESITAVGT VKAMAVEPQMTQRWDNQLAAYVASGFRVT KLAVV 379
 RL++KFAR ADNQSFLVES+TA+ T+KA+AV PQMT WD QLA+YV++GFRVT LA +
 Sbjct: 317 RRLDEKFARGADNQSF LVESTAINTIKALAVTPQMTNTWDKQLASYVSAGFRVTTLATI 376

Query: 380 GQQGVQLIQKLVTVATLWIGARLVIESKLT VQGQLIAFNMLSGQVAPVIRLAQLWQDFQ Q 439
 GQQGVQ IQK+V V TLW+GA LVI L++GQLIAFNMLSGQV APVIRLAQLWQDFQ Q
 Sbjct: 377 GQQGVQFIQKVMVITLWLGAHLVISGDL SIGQLIAFNMLSGQVAPVIRLAQLWQDFQ Q 436

Query: 440 VGISVARLGDILNAPTENASSHLALPDIRGEITFEHVDFRYKADGRLLIQDLNLRIRAGE 499
 VGISV RLGD+LN+PTE+ LALP+I+G+ITF ++ FRYK D +IL D+NL I+ GE
 Sbjct: 437 VGISVTRLGDVLNSPTESYQGLALPEIKGDITFRNIRFRYKPDAPVILNDVNLSIQQGE 496

Query: 500 VLGIVGRSGSGKSTLT KLVRQLYVPAQGRVLVDGNDLALAAPAWLRRQGVV LQENVLLN 559
 V+GIVGRSGSGKSTLT KL+QR Y+P G+VL+DG+DLALA P WLRQGVV LQ+NVLLN
 Sbjct: 497 VIGIVGRSGSGKSTLT KLIRFYIPENGQVLIDGHD LALADPNWLRQGVV LQDNVLLN 556

Query: 560 RSIRDNIALTDTGMPLERIIEAAKLAGAHEFIMELPEGYGT VVGEQAGLSGGQRQRIAI 619
 RSIRDNIAL D GMP+E+I+ AAKLAGAHEFI EL EGY T+VGEQAGLSGGQRQRIAI
 Sbjct: 557 RSIRDNIALADPGMPMEKIVHAAKLAGAHEFISELREGYNTIVGEQAGLSGGQRQRIAI 616

Query: 620 ARALITNPRILIFDEATSALDYESERAIMQNMQAICANRTVLIIAHLSTVKTAHRIIAM 679
 ARAL+ NP+ILIFDEATSALDYESE IM+NM IC RTV+IIAHLSTVK A RII M
 Sbjct: 617 ARALVNNPKILIFDEATSALDYESEHIIMRNMHQICKGRTVIIIAHLSTVK NADRIIVM 676

Query: 680 DKGRIVEAGTQOELLAKPNGYYRYLYDLQ N 709
 +KG+IVE G +ELLA PNG Y YL+ LQ+
 Sbjct: 677 EKGQIVEQGKHKELLADPNGLYHYLHQLQ S 706

Homology with the HlyB leucotoxin secretion ATP-binding protein of *Haemophilus actinomycetemcomitans* (accession number X53955)

ORF39 and HlyB protein show 71% and 69% amino acid identity in 167 and 55 overlap at the N- and C-terminal regions, respectively:

Orf39 1 KFDFTWFIPAVIKYRRXXXXXXXXXXXXXXXXXITPLFFQVMDKVLVHRGFXXXXXXXXX 60
 KFDFTWFIPAVIKYR+ ITPLFFQVMDKVLVHRGF
 HlyB 137 KFDFTWFIPAVIKYRKIFETLIVSIFLQIFALITPLFFQVMDKVLVHRGFSTLNVITV 196

Orf39 61 XXXXXXFEIVLGGRLTYLFAHTTSRIDVELGARLFRHLLSLPLSYFEHRRVGD TVARVR 120
 FEI+LGGRLTY+FAH+TSRIDVELGARLFRHLL+LP+SYFE RRVGD TVARVR
 HlyB 197 ALAIVVLF EII LGGRLTYVFAHSTSRIDVELGARLFRHLLALPISYFEARRVGD TVARVR 256

Orf39 121 ELEQIRNFLT GQALTSVLDLAFSFI FLAVMWYSSSTLTWVVLASLIC 167
 EL+QIRNFLT GQALTS+LDL FSFIF AVMWYYS LT VVL SL C
 HlyB 257 ELDQIRNFLT GQALTSILDLLFSFI FFAVMWYSPKLT LVVLGSLPC 303

//

Orf39 166 ICANRTVLIIAHLSTVKTAHRIIAMDKGRIVEAGTQOELLANXNGYYRYLYDLQ 220
 IC NRTVLIIAHLSTVK A RII MDKG I+E G QELL + G Y YL+ LQ
 HlyB 651 ICQNRTVLIIAHLSTVK NADRIIVMDKGEIIEQGHQELLKDEKGLYSYLHQLQ 705

Based on this analysis, it is predicted that this protein from *N.meningitidis*, and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 7

60 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 31>


```

1  ATGAAATACT TGATCCGCAC CGCCTTACTC GCAGTCGCAG CCGCCGGCAT
51 CTACGCCTGC CAACCGCAAT CCGAAGCCGC AGTGCAAGTC AAGGCTGAAA
101 ACAGCCTGAC CGCTATGCGC TTAGCCGTCG CCGACAAACA GGCAGAGATT
151 GACGGGTTGA ACGCCCAAk sGACCCGAA ATCAGA...

```

5 This corresponds to the amino acid sequence <SEQ ID 32; ORF52>:

```

1  MKYLIRTALL AVAAAGIYAC QPQSEAAVQV KAENSLTAMR LAVADKQAEI
51 DGLNAQXDAE IR..

```

Further work revealed the complete nucleotide sequence <SEQ ID 33>:

```

10      1  ATGAAATACT TGATCCGCAC CGCCTTACTC GCAGTCGCAG CCGCCGGCAT
51      51 CTACGCCTGC CAACCGCAAT CCGAAGCCGC AGTGCAAGTC AAGGCTGAAA
101     101 ACAGCCTGAC CGCTATGCGC TTAGCCGTCG CCGACAAACA GGCAGAGATT
151     151 GACGGGTTGA ACGCCCAAAT CGACGCCGAA ATCAGACAAC GCGAAGCCGA
201     201 AGAATTGAAA GACTACCGAT GGATACACGG CGACGCCGAA GTGCCGGAGC
251     251 TGGAAAAATG A

```

15 This corresponds to the amino acid sequence <SEQ ID 34; ORF52-1>:

```

1  MKYLIRTALL AVAAAGIYAC QPQSEAAVQV KAENSLTAMR LAVADKQAEI
51 DGLNAQIDAE IRQREAEELK DYRWIHGDAE VPELEK*

```

Computer analysis of this amino acid sequence predicts a prokaryotic membrane lipoprotein lipid attachment site (underlined).

20 ORF52-1 (7kDa) was cloned in the pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 4A shows the results of affinity purification of the GST-fusion. Figure 4B shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF52-1.

Based on this analysis, it is predicted that this protein from *N.meningitidis*, and its epitopes, could
25 be useful antigens for vaccines or diagnostics.

Example 8

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 35>

```

30      1  ATGTTATCG GAATATTACT CGCATCAAGC AAGCATGCTC TTGTCATTAC
51      51 TCTATTGTTA AATCCCGTCT TCCATGCATC CAGTTGCGTA TCGCGTTsGG
101     101 CAATACGGAA TAAAtCTGC TGTTCGCTT TGGCTAAATT TGCCAAATTG
151     151 TTTATTGTTT CTTTAGGAGC AGCTTGCTTA GCCGCCTTCG CTTTCGACAA
201     201 CGCCCCACA GGCCTTCCC AAGCgTTGCC TACCGTTACC GCACCCGTGG
251     251 CGATTCCCGC GCCCGCTTCG GCAGCCTGA

```

This corresponds to the amino acid sequence <SEQ ID 36; ORF56>:

```

35      1  MVIGILLASS KHALVITLLL NPVFHASSCV SRXAIRNKIC CSALAKFAKL
51      51 FIVSLGAACL AAFAFDNAPT GASQALPTVT APVAIPAPAS AA*

```

Further work revealed the complete nucleotide sequence <SEQ ID 37>:

```

1  ATGGCTTGTA CAGGTTTGAT GGTTTTCCG TTAATGGTTA TCGGAATATT

```

51 ACTGCATCA AGCAAGCCTG CTCCTTTCCT TACTCTATTG TTAAATCCCG
 101 TCTTCCATGC ATCCAGTTGC GTATCGCGTT GGGCAATACG GAATAAAATC
 151 TGCTGTTCTG CTTTGGCTAA ATTTGCCAAA TTGTTTATTG TTTCTTTAGG
 201 AGCAGCTTGC TTAGCCGCCT TCGCTTTCGA CAACGCCCCC ACAGGCGCTT
 251 CCCAAGCGTT GCCTACCGTT ACCGCACCCG TGGCGATTCC CGCGCCCGCT
 301 TCGGCAGCCT GA

This corresponds to the amino acid sequence <SEQ ID 38; ORF56-1>:

1 MACTGLMVFP LMVIGILLAS SKPAPFLTLL LNPVFHASSC VSRWAI RNKI
 51 CCSALAKFAK LFIVSLGAAC LAAPAFDNAP TGASQALPTV TAPVAIPAPA
 101 SAA*

Computer analysis of this amino acid sequence predicts a leader peptide (underlined) and suggests that ORF56 might be a membrane or periplasmic protein.

Based on this analysis, it is predicted that this protein from *N.meningitidis*, and its epitopes, could be useful antigens for vaccines or diagnostics.

15 Example 9

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 39>

1 ATGTT CAGTA TTTTAAATGT GTTTCTTCAT TGTATTCTGG CTTGTGTAGT
 51 CTC TGGTGAG ACGCCTACTA TATTTGGTAT CCTTGCTCTT TTTTACTTAT
 101 TGTATCTTTC TTATCTTGCT GTTTTAAAGA TTTTCTTTTC TTTTCTTCTA
 151 GACAGAGTTT CACTCCGGTC TCCCAGGCTG GAGTGCAAT GGCATGACCC
 201 TTTGGCTCAC TGGCTCACGG CCACCTCTGC TATTCTGCCG CCTCAGCCTC
 251 CAGGG...

This corresponds to the amino acid sequence <SEQ ID 40; ORF63>:

1 MFSILNVFLH CILACVVSGE TPTIFGILAL FYLLYLSYLA VFKEIFFSFFL
 251 DRVSLRSPRL ECKWHDPLAH WLTATSAILP PQPPG...

Computer analysis of this amino acid sequence predicts a transmembrane region.

Based on this analysis, it is predicted that this protein from *N.meningitidis*, and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 10

30 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 41>

1 ..GTGCGGACGT GGTGGTTTT TTGGTTGCAG CGTTTGAAAT ACCCGTTGTT
 51 GCTTTGGATT GCGGATATGT TGCTGTACCG GTTGTGGGC GGCGCGGAAA
 101 TCGAATGCGG CCGTTGCCCT GTGCGGCCGA TGACGGATTG GCAGCATTTT
 151 TTGCCGGCGA TGGGAACGGT GTCGGCTTGG GTGGCGGTGA TTTGGGCATA
 201 CCTGATGATT GAAAGTGAAA AAAACGGAAG ATATTGA

This corresponds to the amino acid sequence <SEQ ID 42; ORF69>:

1 ..VRTWLVFVWLQ RLKYPLLLWI ADMLLYRLLG GAEIECGRCP VPPMTDWQHF
 51 LPAMGTVSAW VAVIWAYLMI ESEKNGRY*

Computer analysis of this amino acid sequence predicts a transmembrane region.

A corresponding ORF from strain A of *N.meningitidis* was also identified:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF69 shows 96.2% identity over a 78aa overlap with an ORF (ORF69a) from strain A of *N.*

meningitidis:

		10	20	30	40	50	60
orf69.pep		VRTWL VFWLQRLKY PLLLWIADMLLYRL LGGAEIECGRCPVPPMTDWQHFLPAMGT VSAW					
orf69a		VRTWL VFWLQRLKY PLLLWIADMLLYRL LGGAEIECGRCPVPPMTDWQHFLPTMGTVAAW					
		10	20	30	40	50	60
		70	79				
orf69.pep		VAVIWAYLMIESEKNGRYX					
orf69a		VAVIWAYLMIESEKNGRYX					
		70					

The ORF69a nucleotide sequence <SEQ ID 43> is:

1	GTGCGGACGT	GGTTGGTTT	TTGGTTGCAG	CGTTTGAAT	ACCCGTTGTT
51	GCTTTGTATT	GCGGATATGC	TGCTGTACCG	GTTGTTGGGC	GGCGCGGAAA
101	TCGAATGCGG	CCGTTGCCCT	GTACCGCCGA	TGACGGATTG	GCAGCATTTT
151	TTGCCGACGA	TGGGAACGGT	GGCGGCTTGG	GTGGCGGTGA	TTTGGGCATA
201	CCTGATGATT	GAAAGTGAAA	AAAACGGAAG	ATATTGA	

This encodes a protein having amino acid sequence <SEQ ID 44>:

1	VRTWL VFWLQ	RLKY PLLLCI	ADMLLYRL LG	GAEIECGRCP	VPPMTDWQHF
51	LPTMGTVAAW	VAVIWAYLMI	ESEKNGRY*		

Based on this analysis, it is predicted that this protein from *N.meningitidis*, and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 11

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 45>

1	ATGTTTCAAA	ATTTTGATT	GGGCGTGTTC	CTGCTTGCCG	TCCTCCCCGT
51	GCTGCCCTCC	ATTACCGTCT	CGCACGTGGC	GCGCGGCTAT	ACGGCGCGCT
101	ACTGGGGAGA	CAACACTGCC	GAACAATACG	GCAGGCTGAC	ACTGAACCCC
151	CTGCCCCATA	TCGATTGGT	CGGCACAATC	ATCgTACCGC	TGCTTACTTT
201	GATGTTACAG	CCCTTCCTGT	TCGGCTGGGC	GCGTCCGATT	CCTATCGATT
251	CGCGCAACTT	CCGCAACCCG	cGCCTTGCC	GGCGTTGCGT	TGCCGCGTCC
301	GGCCCGCTGT	CGAATCTAGC	GATGGCTGTW	CTGTGGGGCG	TGGTTTGGT
351	GCTGACTCCG	TATGTCGGCG	GGGCGTATCA	GATGCCGTTG	GCTCAAATGG
401	CAAACTACGG	TATCTGATC	AATGCGATTG	TGTTCGCGCT	CAACATCATC
451	CCCATCCTGC	CTTGGGACGG	CGGCATTTTC	ATCGACACCT	TCCTGTCGGC
501	GAAATATTTC	CAAGCGTTCC	GCAAAATCGA	ACCTTATGGG	ACGTGGATTA
551	TCCTACTGCT	GATGCTGACC	sGGGTTTGG	GTGCGTTTAT	wGCACCGATT
601	sTGCGGmTGc	GTGATTGCrT	TTGTGCAGAT	GTWCGTCTGA	CTGGCTTTCA
651	GACGGCATAA				

This corresponds to the amino acid sequence <SEQ ID 46; ORF77>:

1 MFQNFDLGVF LLAVLPVLPS ITVSHVARGY TARYWGDNTA EQYGRLLNLP
 51 LPHIDLVGTI IVPLLLTMFT PFLFGWARPI PIDSRNFRNP RLAWRCVAAS
 101 GPLSNLAMAV LWGVVLVLTPL YVGGAYQMP L AQMANYGILI NAILFALNII
 151 PILPWDGGIF IDTFLSAKYS QAFRKIEPYG TWIILLMLT XVLGAFTAPI
 201 XRXRDCXCAD VRLTGFQTA*

Further work revealed the complete nucleotide sequence <SEQ ID 47>:

1 ATGTTTCAAA ATTTTGATTT GGGCGTGTTC CTGCTTGCCG TCCTGCCCGT
 51 GCTGCTCTCC ATTACCGTCA GGGAGGTGGC GCGCGGCTAT ACGGCGCGCT
 101 ACTGGGGAGA CAACACTGCC GAACAATACG GCAGGCTGAC ACTGAACCCC
 151 CTGCCCCATA TCGATTGGT CCGCACATC ATCGTACCGC TGCTTACTTT
 201 GATGTTACAG CCCTTCCTGT TCGGCTGGG CCGTCCGATT CCTATCGATT
 251 CGCGCAACTT CCGCAACCCG CGCCTTGCTT GCGGTTGCGT TGCCGCGTCC
 301 GGCCCGCTGT CGAATCTAGC GATGGCTGTT CTGTGGGGCG TGGTTTGGT
 351 GCTGACTCCG TATGTCGGCG GGGCGTATCA GATGCCGTG GCTCAAATGG
 401 CAAACTACGG TATTCTGATC AATGCGATT TGTTCGCGCT CAACATCATC
 451 CCCATCCTGC CTTGGGACGG CGGCATTTTC ATCGACACCT TCCTGTCCGG
 501 GAAATATTCG CAAGCGTTCC GCAAAATCGA ACCTTATGGG ACGTGGATTA
 551 TCCTACTGCT GATGCTGACC GGGGTTTGG GTGCGTTTAT TGCACCGATT
 601 GTGCGGCTGG TGATTGCGTT TGTGCAGATG TTCGTCTGA

This corresponds to the amino acid sequence <SEQ ID 48; ORF77-1>:

1 MFQNFDLGVF LLAVLPVLLS ITVREVARGY TARYWGDNTA EQYGRLLNLP
 51 LPHIDLVGTI IVPLLLTMFT PFLFGWARPI PIDSRNFRNP RLAWRCVAAS
 101 GPLSNLAMAV LWGVVLVLTPL YVGGAYQMP L AQMANYGILI NAILFALNII
 151 PILPWDGGIF IDTFLSAKYS QAFRKIEPYG TWIILLMLT GVLGAFTAPI
 201 VRLVIAFVQM FV*

Computer analysis of this amino acid sequence reveals a putative leader sequence and several transmembrane domains.

A corresponding ORF from strain A of *N.meningitidis* was also identified:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF77 shows 96.5% identity over a 173aa overlap with an ORF (ORF77a) from strain A of *N.meningitidis*:

		10	20	30	40	50	60
orf77.pep	MFQNFDLGVFLLAVLPVLPSITVSHVARGYTARYWGDNTAEQYGRLLTNLPHIDLVGTI						
orf77a	RGYTARYWGDNTAEQYGRLLTNLPHIDLVGTI						
		10	20	30			
		70	80	90	100	110	120
orf77.pep	IVPLLLTMFTPFLFGWARPIPIDSRNFRNPRLAWRCVAASGPLSNLAMAVLWGVVLVLTPL						
orf77a	IVPLLLTMFTPFLFGWARPIPIDSRNFRNPRLAWRCVAASGPLSNLAMAVLWGVVLVLTPL						
		40	50	60	70	80	90
		130	140	150	160	170	180
orf77.pep	YVGGAYQMP LAQMANYGILINAILFALNIIPILEPWDGGIFIDTFLSAKYSQAFRKIEPYG						
orf77a	YVGGAYQMP LAQMANYXILINAILXALNIIPILEPWDGGIFIDTFLSAKXSQAFRKIEPYG						
		100	110	120	130	140	150
		190	200	210	220		
orf77.pep	TWIILLMLTXVLGAFTAPIXRXRDCXCADVRLTGFQTA						
orf77a	TWIIKLLMLTGVLGAXIAPIVQLVIAFVQMFVX						
		160	170	180			

ORF77-1 and ORF77a show 96.8% identity in 185 aa overlap:

		10	20	30	40	50	60
5	orf77-1.pep	MFQNF	DLGVFLLAVLPVLLSITVRE	VARGYTARYWGDNTAEQYGR	LTNLPLPHIDL	VGTI	
	orf77a			RGYTARYWGDNTAEQYGR	LTNLPLPHIDL	VGTI	
				10	20	30	
10	orf77-1.pep	IVPLL	TLMFTFPLFGWARPI	PIDSRNERNPRLAWRCVAASG	PLSNLAMAVLWGV	VVLVLT	P
	orf77a	IVPLL	TLMFTFPLFGWARPI	PIDSRNERNPRLAWRCVAASG	PLSNLAMAVLWGV	VVLVLT	P
		40	50	60	70	80	90
15	orf77-1.pep	YVGGAYQ	MPLAQMANYGILINAILFAL	NIIPILPWDGGIFIDTFL	SAKYSQAFRKIE	PYG	
	orf77a	YVGGAYQ	MPLAQMANYXILINAILXAL	NIIPILPWDGGIFIDTFL	SAKXSQAFRKIE	PYG	
		100	110	120	130	140	150
20	orf77-1.pep	TWIIILL	MLTGVLGAFIPIVRLVIA	FVQMFVX			
	orf77a	TWIIIXL	MLTGVLGAXIPIVQLVIA	FVQMFVX			
25		160	170	180			

A partial ORF77a nucleotide sequence <SEQ ID 49> was identified:

	1	..CGCGGCTATA	CAGCGCGCTA	CTGGGGTGAC	AACACTGCCG	AACAATACGG
	51	CAGGCTGACA	CTGAACCCCC	TGCCCCATAT	CGATTGGTC	GGCACAATCA
30	101	TCGTACCGCT	GCTTACTTTG	ATGTTTACGC	CCTTCCTGTT	CGGCTGGGCG
	151	CGTCCGATT	CTATCGATT	GCGCAACTTC	CGCAACCCGC	GCCTTGCTG
	201	GCGTTGCGTT	GCCGCGTCCG	GCCCGCTGTC	GAATCTGGCG	ATGGCTGTT
	251	TGTGGGGCGT	GGTTTGGTG	CTGACTCCGT	ATGTCGGTGG	GGCGTATCAG
	301	ATGCCGTTGG	CNCAATGGC	AAACTACNNN	ATTCTGATCA	ATGCGATTCT
	351	GTNCGCGCTC	AACATCATCC	CCATCCTGCC	TTGGGACGGC	GGCATTTTCA
35	401	TCGACACCTT	CCTGTCGGCN	AAATANTCGC	AAGCGTCCG	CAAAATCGAA
	451	CCTTATGGGA	CGTGGAATTA	CCNGCTGCTT	ATGCTGACCG	GGGTTTGGG
	501	TGCGTNTATT	GCACCGATTG	TGCAGCTGGT	GATTGCGTTT	GTGCAGATGT
	551	TCGTCCTGA				

This encodes a protein having amino acid sequence <SEQ ID 50>:

40	1	..RGYTARYWGD	NTAEQYGR	LTNLPLPHIDL	VGTIIVPLLTL	MFTPFLGWA
	51	RPIPIDSRNF	RNPRLAWRCV	AASGPLSNLA	MAVLWGVVLV	LPYVGGAYQ
	101	MPLAQMANYX	ILINAILXAL	NIIPILPWDG	GIFIDTFLSA	KXSQAFRKIE
	151	PYGTWIIIXLL	MLTGVLGAXI	APIVQLVIAF	VQMFV*	

Based on this analysis, it is predicted that this protein from *N.meningitidis*, and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 12

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 51>

	1	ATGAACCTGA	TTTACAGTTA	CATCATCCGT	CAAATGGCGG	TTATGGCGGT
	51	TTACGCGCTC	CTTGCCTTCC	TCGCTTTGTA	CAGCTTTTTT	GAAATCCTGT
50	101	ACGAAACCGG	CAACCTCGGC	AAAGGCAGTT	ACGGCATATG	GGAAATGCTG
	151	GGCTACACCG	CCCTCAAAAT	GCCCGCCCGC	GCCTACGAAC	TGATTCCCCT
	201	CGCCGTCCTT	ATCGGCGGAC	TGGTCTCCCT	CAGCCAGCTT	GCCGCCGGCA
	251	GCGAACTGAC	CGTCATCAAA	GCCAGCGGCA	TGAGCACCAA	AAAGCTGCTG
	301	TTGATTCTGT	CGCAGTTCGG	TTTTATTITT	GCTATTGCCA	CCGTCGCGCT
55	351	CGCGGAATGG	GTTGCGCCCA	CACTGAGCCA	AAAAGCCGAA	AACATCAAAG

401 CCGCCGCCAT CAACGGCAAA ATCAGCACCG GCAATACCGG CCTTTGGCTG
451 AAAGAAAAAA ACAGCGTGAT CAATGTGCGC GAAATGTTGC CCGACCAT..

This corresponds to the amino acid sequence <SEQ ID 52; ORF112>:

1 MNLSRYIIR QMAVMAVYAL LAFLALYSFF EILYETGNLG KGSYGIWEML
51 GYTALKMPAR AYELIPLAVL IGGLVLSLSQL AAGSELTVIK ASGMSTKKLL
101 LILSQFGFIF AIATVALGEW VAPTLSQKAE NIKAAAINGK ISTGNTGLWL
151 KEKNSVINVR EMLPDH...

Further work revealed further partial nucleotide sequence <SEQ ID 53>:

1 ATGAACCTGA TTTCACGTTA CATCATCCGT CAAATGGCGG TTATGGCGGT
51 TTACGCGCTC CTTCGCTTCC TCGCTTTGTA CAGCTTTTTC GAAATCCTGT
101 ACGAAACCGG CAACCTCGGC AAAGGCAGTT ACGGCATATG GGAATGCTG
151 gGCTACACCG CCTCAAAT GCGCGCCGCG GCCTACGAAC TGATTCCCT
201 CGCGCTCCTT ATCGGCGGAC TGGTCTCCCT CAGCCAGCTT GCGCGCGCA
251 GCGAACTGAC CGTCATCAAA GCCAGCGGCA TGAGCACCAA AAAGCTGCTG
15 301 TTGATTCTGT CGCAGTTCGG TTTTATTTT GCTATTGCCA CCGTCGCGCT
351 CGGCGAATGG GTTGCGCCA CACTGAGCCA AAAAGCCGAA AACATCAAG
401 CGCGCGCAT CAACGGCAAA ATCAGCACCG GCAATACCGG CCTTTGGCTG
451 AAAGAAAAAA ACAGCTKAT CAATGTGCGC GAAATGTTGC CCGACCATAC
501 GCTTTTGGGC ATCAAAATTT GGGCGCGCAA CGATAAAAC GAATTGGCAG
20 551 AGGCAGTGA AGCCGATTCC GCCGTTTGA ACAGCGACGG CAGTTGGCAG
601 TTGAAAAACA TCCGCCGCGC CACGCTTGGC GAAGACAAAG TCGAGGTCTC
651 TATTGCGGCT GAAGAAACT GGCCGATTTC CGTCAAACGC AACCTGATGG
701 ACGTATTGCT CGTCAAACCC GACCAAATGT CCGTCGGCGA ACTGACCACC
751 TACATCCGCC ACCTCCAAA CAACAGCCAA AACACCCGAA TCTACGCCAT
25 801 CGCATGGTGG CGCAAATTGG TTACCCCGC CGCAGCCTGG GTGATGGCGC
851 TCGTCGCCTT TGCCTTTACC CCGCAAACCA CCGCCACGG CAATATGGGC
901 TTAATACTCT TCGCGGCAT CTGTsTCGGA TTGCTGTTC ACCTTCCGGG
951 ACGGCTCTTT GGGTTTACCA GCGAACTCGG...

This corresponds to the amino acid sequence <SEQ ID 54; ORF112-1>:

1 MNLSRYIIR QMAVMAVYAL LAFLALYSFF EILYETGNLG KGSYGIWEML
51 GYTALKMPAR AYELIPLAVL IGGLVLSLSQL AAGSELTVIK ASGMSTKKLL
101 LILSQFGFIF AIATVALGEW VAPTLSQKAE NIKAAAINGK ISTGNTGLWL
151 KEKNSXINVR EMLPDHTLLG IKIWARNDKN ELAEAVEADS AVLNSDGSWQ
201 LKNIRRTLK EDKVEVSIAA EENWPISVKR NIMDVLLVKP DOMSVGELTT
251 YIRHLQNSQ NTRIYAIWW RKLVPAAAW VMLVAFAFT PQTTRHNGMG
301 LKLFGGICKG LLFHLAGRLF GFTSQL...

Computer analysis of this amino acid sequence predicts two transmembrane domains.

A corresponding ORF from strain A of *N.meningitidis* was also identified:

Homology with a predicted ORF from *N.meningitidis* (strain A)

40 ORF112 shows 96.4% identity over a 166aa overlap with an ORF (ORF112a) from strain A of *N.meningitidis*:

		10	20	30	40	50	60
45	orf112.pep	MNLSRYIIRQMAVMAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMLGYTALKMPAR					
	orf112a	MNLSRYIIRQMAVMAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMXGYTALKMXAR					
		10	20	30	40	50	60
50	orf112.pep	AYELIPLAVLIGGLVLSLSQLAAGSELTVIKASGMSTKKLLLILSQFGFIFAIATVALGEW					
	orf112a	AYELMPLAVLIGGLVXSQLAAGSELXVIKASGMSTKKLLLILSQFGFIFAIATVALGEW					
		70	80	90	100	110	120

```

      130      140      150      160
orf112.pep VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKNSVINREMLPDH
            |||||:|||||
orf112a    VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKNSIINVREMLPDHTLLGIKIWARNDKN
      130      140      150      160      170      180
orf112a    ELAEAVEADSAVLNSDGSWQLKNIRRLSTLGEDKVEVSIAAEEXWPISVKRNLMDVLLVKP
            190      200      210      220      230      240

```

A partial ORF112a nucleotide sequence <SEQ ID 55> was identified:

```

10      1  ATGAACCTGA  TTTCACGTTA  CATCATCCGT  CAAATGGCGG  TTATGGCGGT
      51  TTACGCGCTC  CTTGCCTTCC  TCGCTTTGTA  CAGCTTTTTT  GAAATCCTGT
     101  ACGAAACCGG  CAACCTCGGC  AAAGGCAGTT  ACGGCATATG  GGAATGNTG
     151  GGNATACACG  CCTCAAAAT  GNCCGCCCGC  GCCTACGAAC  TGATGCCCT
     201  CGCGTCTCTT  ATCGGCGGAC  TGGTCTCTNT  CAGCCAGCTT  GCCGCCGGCA
     251  GCGAACTGAN  CGTCATCAA  GCCAGCGGCA  TGAGCACCAA  AAAGCTGCTG
     301  TTGATTCTGT  CGCAGTTCGG  TTTTATTTT  GCTATTGCCA  CCGTCGCGCT
     351  CGGCGAATGG  GTTGCGCCCA  CACTGAGCCA  AAAAGCCGAA  AACATCAAG
     401  CCGCGGCCAT  CAACGGCAA  ATCAGTACCG  GCAATACCG  CCTTTGGCTG
     451  AAAGAAAAAA  ACAGCATTAT  CAATGTGCGC  GAAATGTTGC  CCGACCATAC
     501  CCGTCTGGGC  ATTAAAATCT  GGGCCCGCAA  CGATAAAAC  GAACTGGCAG
     551  AGGCAGTGG  AGCCGATTCC  GCCGTTTGA  ACAGCGACGG  CAGTTGGCAG
     601  TTGAAAAACA  TCCGCCGCG  CACGCTTGGC  GAAGACAAAG  TCGAGGTCTC
     651  TATTGCGGCT  GAAGAAAANT  GGCCGATTTC  CGTCAAACGC  AACCTGATGG
     701  ACGTATTGCT  CGTCAAACCC  GACCAAATGT  CCGTCGGCG  ACTGACCACC
     751  TACATCCGCC  ACCTCCAAAN  NNACAGCCAA  AACACCGGAA  TCTACGCCAT
     801  CGCATGGTGG  CGCAAATTGG  TTTACCCCGC  CGCAGCCTGG  GTGATGGCGC
     851  TCGTCGCCCT  TGCCTTTACC  CCGCAAACCA  CCCGCCACGG  CAATATGGGC
     901  TTAAAANTCT  TCGCGGCGAT  CTGTCTCGGA  TTGTCTGTCC  ACCTTGCCGG
     951  NCGGCTCTTC  NGGTTTACCA  GCCAACTCTA  CGGCATCCCG  CCCTTCCTCG
    1001  NCGGCGCACT  ACCTACCATA  GCCTTCGCCT  TGCTCGCCGT  TTGGCTGATA
    1051  CGCAACAGG  AAAACGCTA  A

```

This encodes a protein having amino acid sequence <SEQ ID 56>:

```

      1  MNLSRYIIR  QMAVMAVYAL  LAFLALYSFF  EILYETGNLG  KGSYGIWEMX
     51  GYTALKMXAR  AYELMPLAVL  IGGLVXSXSL  AAGSELXVIK  ASGMSTKLL
    101  LILSQFGFIF  AIATVALGEW  VAPTLSQKAE  NIKAAAINGK  JGNTGLWL
    151  KEKNSIINVR  EMLPDHTLLG  IKIWARNDKN  ELAEAVEADS  AVLNSDGSWQ
    201  LKNIRRLSTL  EDKVEVSIAA  EEXWPISVKR  NLMDVLLVKP  DQMSVGELTT
    251  YIRHLQXXSQ  NTRIYAIAWW  RKLVPAAAW  VMLVAFAPT  PQTTRHGNMG
    301  LKXFGGICLG  LLFHLAARLF  XFTSQLYGIP  PFLXGALPTI  AFALLAVWLI
    351  RKQEKRR*

```

ORF112a and ORF112-1 show 96.3% identity in 326 aa overlap:

```

      orf112a.pep  MNLSRYIIRQMAVMAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMXGYTALKMXAR
      |||||
      orf112-1    MNLSRYIIRQMAVMAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMLGYTALKMPAR
45      orf112a.pep  AYELMPLAVLIGGLVXSXSLAAGSELXVIKASGMSTKLLLLILSQFGFIFAIATVALGEW
      |||||
      orf112-1    AYELIPLAVLIGGLVLSQLAAGSELTVIKASGMSTKLLLLILSQFGFIFAIATVALGEW
50      orf112a.pep  VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKNSIINVREMLPDHTLLGIKIWARNDKN
      |||||
      orf112-1    VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKNSXINVREMLPDHTLLGIKIWARNDKN
55      orf112a.pep  ELAEAVEADSAVLNSDGSWQLKNIRRLSTLGEDKVEVSIAAEEXWPISVKRNLMDVLLVKP
      |||||
      orf112-1    ELAEAVEADSAVLNSDGSWQLKNIRRLSTLGEDKVEVSIAAEENWPISVKRNLMDVLLVKP
60      orf112a.pep  DQMSVGELTTYIRHLQXXSQNTRIYAIAWWRKLVPAAAWVMLVAFAPTQTTRHGNMG
      |||||
      orf112-1    DQMSVGELTTYIRHLQNNQNTRIYAIAWWRKLVPAAAWVMLVAFAPTQTTRHGNMG
      orf112a.pep  LKXFGGICLGLLFLHLAGRLEFXFTSQLYGIPPFLXGALPTIAFALLAVWLIRKQEKRR
      || |||||

```

orf112-1

LKLFGGICXGLLFHLAGRLFGFTSQL

Based on this analysis, it is predicted that this protein from *N.meningitidis*, and its epitopes, could be useful antigens for vaccines or diagnostics.

5 Example 13

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 57>

```

1  ..GCAGTAGCCG AAACTGCCAA CAGCCAGGGC AAAGGTAAAC AGGCAGGCAG
51  TTCGGTTTCT GTTTCCTGA AACTTCAGG CGACCTTTGC GGCAAACTCA
101 AAACCACCCT TAAAACTTTG GTCTGCTCTT TGGTTTCCCT GAGTATGGTA
151 TTGCCTGCCC ATGCCCAAAT TACCACGAC AAATCAGCAC CTAAAAACCA
201 GCAGGTCGTT ATCCTTAAAA CCAACACTGG TGCCCTTTG GTGAATATCC
251 AAACTCCGAA TGGACGCGGA TTGAGCCACA ACCGCTA.TA CGCATTTGAT
301 GTTGACAACA AAGGGGCACT GTTAAACAAC GACCGTAACA ATAATCCGTT
351 TGTGGTCAAA GGCAGTGCGC AATTGATTTT GAACGAGGTA CGCGGTACGG
15 401 CTAGCAAACCT CAACGGCATC GTTACCGTAG GCGGTCAAAA GGCGACGTG
451 ATTATTGCCA ACCCAACCGG CATTACCGTT AATGGCGGCG GCTTTAAAAA
501 TGTCGGTCCG GGCATCTTAA CTACCGGTGC GCCCAAATC GGCAAGACG
551 GTGCACTGAC AGGATTTGAT GTGCGTCAAG GCACATTGGA CCGTAGrAGC
601 AGCAGGTTGG AATGATAAAG GCGGAGCmrm yTACACCGGG GTACTTGCTC
20 651 GTGCAGTTGC TTTGCAGGGG AAATTwmrGG GTAAA.AACT GGCGGTTTCT
701 ACCGGTCCTC AGAAAGTAGA TTACGCCAGC GCGGAAATCA GTGCAGGTAC
751 GCGAGCGGGT ACGAAACCGA CTATTGCCCT TGATACTGCC GCACCTGGCG
801 GTATGTACGC CGACAGCATC AACTGATTG CCAATGAAAA AGGCGTAGGC
851 GTCTAA

```

25 This corresponds to the amino acid sequence <SEQ ID 58; ORF114>:

```

1  ..AVAETANSQG KGKQAGSSVS VSLKTSGLDLC GKLKTTLKTLC VCSLVSLSMV
51  LPAHAQITTD KSAPKNQOVV ILKTNLTGAPL VNIQTPNGRG LSHNRXYAFD
101 VDNKGAVLNN DRNNNPFVVK GSAQLILNEV RGTASKLNGI VTVGQKADV
151 IIANPNGITV NGGGFKNVGR GILTTGAPQI GKDGALTGFD VVKAHWTvYA
30 201 AGWNDKGGAX YTGVLARAVA LQKXXGKXL AVSTGPQKVD YASGEISAGT
251 AAGTKPTIAL DTAALGGMYA DSITLIANEK GVGv*

```

Further work revealed the complete nucleotide sequence <SEQ ID 59>:

```

1  ATGAATAAAG GTTTACATCG CATTATCTTT AGTAAAAAGC ACAGCACCAT
35 51  GGTTCAGTA GCCGAACTG CCAACAGCCA GGGCAAAGGT AAACAGGCAG
101 GCAGTTCGGT TTCTGTTTCA CTGAAACTT CAGGCGACCT TTGCGGCAAA
151 CTCAAAACCA CCCTTAAAC TTTGGTCTGC TCTTTGGTTT CCCTGAGTAT
201 GGTATTGCCT GCCCATGCCC AAATTACCAC CGACAAATCA GCACCTAAAA
251 ACCAGCAGGT CGTTATCCTT AAAACCAACA CTGGTGCCCC CTTGGTGAAT
301 ATCCAAACTC CGAATGGACG CGGATTGAGC CACAACCGCT ATACGCAGTT
40 351 TGATGTTGAC AACAAAGGGG CAGTGTAAA CAACGACCGT AACAATAATC
401 CGTTTGTGGT CAAAGGCACT GCGCAATTGA TTTGAACGA GGTACGCGGT
451 ACGGCTAGCA AACTCAACGG CATCGTTACC GTAGGCGGTC AAAAGGCCGA
501 CGTGATTATT GCCAACCCCA ACGGCATTAC CGTTAATGGC GGCGGCTTTA
551 AAAATGTCGG TCGGGGCATC TTAACTACCG GTGCGCCCCA AATCGGCAAA
45 601 GACGGTGCAC TGACAGGATT TGATGTGCGT CAAGGCACAT TGACCGTAGG
651 AGCAGCAGGT TGAATGATA AAGGCGGAGC CGACTACACC GGGGTACTTG
701 CTCGTGCAGT TGCTTTGCAG GGGAAATTAC AGGGTAAAAA CCTGGCGGTT
751 TCTACCGGTC CTCAGAAAGT AGATTACGCC AGCGGCGAAA TCAGTGCAGG
50 801 TACGGCAGCG GTACGAAAC CGACTATTGC CCTTGATACT GCCGCACTGG
851 GCGGTATGTA CGCCGACAGC ATCACAATGA TTGCCAATGA AAAAGGCGTA
901 GCGGTCAAAA ATGCGGCGAC ACTCGAAGCG GCCAAGCAAT TGATTGTGAC
951 TTCGTGAGG CGCATTGAAA ACAGCGGCGG CATCGCCACC ACTGCCGAGC
1001 GCACCGAAGC TTCACCGACT TATCTCTCCA TCGAAACCAAC CGAAAAAGGA
1051 GCGGCAGGCA CATTTATCTC CAATGGTGGT CGGATCGAGA GCAAGGCTT
55 1101 ATTGGTTATT GAGACGGGAG AAGATATCAG CTGCGTAAAC GGAGCCGTGG
1151 TGCAGATAAA CGGCAGTCGC CCAGTACCA CGGTATTAAA TGCTGGTCAT
1201 AATTTGGTGA TTGAGAGCAA AACTAATGTG AACAATGCCA AAGGCCCGGC

```


1251 TACTCTGTCCG GCCGACGGCC GTACCGTCAT CAAGGAGGCC AGTATTGAGA
1301 CTGGCACTAC CGTATACAGT TCCAGCAAAG GCAACGCCGA ATTAGGCAAT
1351 AACACACGCA TTACCGGGGC AGATGTTACC GTATTATCCA ACGGCACCAT
1401 CAGCAGTTCC GCCGTAATAG ATGCCAAAGA CACCGCACAC ATCGAAGCAG
1451 GCAAAACGCT TTCTTTGGAA GCTTCAACAG TTACCTCCGA TATCCGCTTA
1501 AACGGAGGCA GTATCAAGGG CGGCAAGCAG CTTGCTTTAC TGGCAGACGA
1551 TAACATTACT GCCAAAATA CCAATCTGAA TACTCCCGGC AATCTGTATG
1601 TTCATACAGG TAAAGATCTG AATTGGAATG TTGATAAAGA TTTGTCTGCC
1651 GCCAGCATCC ATTTGAAATC GGATAACGCT GCCCATATTA CCGGCACCAG
1701 TAAAACCCCTC ACTGCCTCAA AAGACATGGG TGTGGAGGCA GGCTCGCTGA
1751 ATGTTACCAA TACCAATCTG CGTACCAACT CGGGTAATCT GCACATTCAG
1801 GCAGCCAAAG GCAATATTCA GCTTCGCAAT ACCAAGCTGA ACGCAGCCAA
1851 GGCTCTCGAA ACCACGCGAT TGCAGGGCAA TATCGTTTCA GACGGCCTTC
1901 ATGCTGTTTC TGCAGACGGT CATGTATCCT TATTGGCCAA CGGTAATGCC
1951 GACTTTACCG GTCACAATAC CCTGACAGCC AAGGCCGATG TCAATGCAGG
2001 ATCGGTTGGT AAAGGCCGTC TGAAAGCAGA CAATACCAAT ATCACTTCAT
2051 CTTCAGGAGA TATTACGTTG GTTGCCGGCA ACGGTATTCA GCTTGGTGAC
2101 GGAAAACAAC GCAATTCAAT CAACGGAAAA CACATCAGCA TCAAAAACAA
2151 CGGTGGTAAT GCCGACTTAA AAAACCTTAA CGTCCATGCC AAAAGCGGGG
2201 CATTGAACAT TCATTCCGAC CGGGCATTGA GCATAGAAAA TACCAAGCTG
2251 GAGTCTACCC ATAATACGCA TCTTAATGCA CAACACGAGC GGGTAACGCT
2301 CAACCAAGTA GATGCCTACG CACACCGTCA TCTAAGCATT ACCGGCAGCC
2351 AGATTGGCA AAACGACAAA CTGCCCTTCTG CCAACAAGCT GGTGGCTAAC
2401 GGTGTATTGG CACTCAATGC GCGCTATTCC CAAATTGCCG ACAACACCAC
2451 GCTGAGAGCG GGTGCAATCA ACCTTACTGC CGGTACCGCC CTAGTCAAGC
2501 GCGGCAACAT CAATTGGAGT ACCGTTTCGA CCAAACTTT GGAAGATAAT
2551 GCCGAATTAA AACCATGGC CGGACGGCTG AATATTGAAG CAGGTAGCGG
2601 CACATTAAAC ATCGAACCTG CCAACCGCAT CAGTGCGCAT ACCGACCTGA
2651 GCATCAAAAC AGGCGAAAA TTGCTGTTGT CTGCAAAAGG AGGAAATGCA
2701 GGTGCGCCTA GTGCTCAAGT TTCTCTATTG GAAGCAAAAG GCAATATCCG
2751 TCTGGTTACA GGAGAAACAG ATTTAAGAGG TTCTAAATTT ACAGCCGGTA
2801 AAAACTTGGT TGTGCGCCACC ACCAAAGGCA AGTTGAATAT CGAAGCCGTA
2851 AACAACTCAT TCAGCAATTA TTTTCTTACA CAAAAGCGG CTGAATCTCA
2901 CCAAAAATCC AAAGAATTGG AACAGCAGAT TGCGCAGTTG AAAAAAGCT
2951 CGCCTAAAG CAAGCTGATT CCAACCTGCG AAGAAGAACG CGACCGTCTC
3001 GCTTTCTATA TTCAAGCCAT CAACAAGGAA GTTAAAGGTA AAAAACCCAA
3051 AGGCAAGAA TACCTGCAAG CCAAGCTTTC TGCACAAAAT ATTGACTTGA
3101 TTTCCGCACA AGGCATCGAA ATCAGCGGTT CCGATATTAC CGCTTCCAAA
3151 AAAGTGAACC TTCACGCCGC AGGCGTATTG CCAAGGCAG CAGATTGAGA
3201 GGCGGCTGCT ATTCTGATTG ACGGCATAAC CGACCAATAT GAAATTGGCA
3251 AGCCCACTA CAAGAGTCAC TACGACAAAG CTGCTCTGAA CAAGCCTTCA
3301 CGTTTGACCG GACGTACAGG GGTAAGTATT CATGCAGCTG CGGCACTCGA
3351 TGATGCACGT ATTATTATCG GTGCATCCGA AATCAAAGCT CCCTCAGGCA
3401 GCATAGACAT CAAAGCCCAT AGTGATATTG TACTGGAGGC TGGACAAAAC
3451 GATGCTTATA CCTTCTTAAA AACCAAAGGT AAAAGCGGCA AAATCATCAG
3501 AAAAACCAAG TTTACAGCA CCCGCGACCA CTGATTATG CCAGCCCCCG
3551 TCGAGCTGAC CGCCAACGGC ATAACGCTTC AGGCAGGCGG CAACATCGAA
3601 GCTAATACCA CCCGCTTCAA TGCCCCTGCA GGTAAAGTTA CCCTGGTTGC
3651 GGGTGAAGAG CTGCAACTGC TGGCAGAAGA AGGCATCCAC AAGCACGAGT
3701 TGGATGTCCA AAAAAGCCGC CGCTTTATCG GCATCAAGGT AGGCAAGAGC
3751 AATTACAGTA AAAACGAACT GAACGAAACC AAATTGCCTG TCCGCGTCGT
3801 CGCCCAAACT GCAGCCACCC GTTCAGGCTG GGATACCGTG CTCGAAGGTA
3851 CCGAATTCAA AACCACGCTG GCCGGTGCGG ACATTACGGC AGGTGTAGGC
3901 GAAAAAGCCC GTGCCGATGC GAAAATTATC CTCAAAGGCA TTGTGAACCG
3951 TATCCAGTCG GAAGAAAAAT TAGAAACCAA CTCAACCGTA TGGCAGAAAC
4001 AGGCCGGACG CGGCAGCACT ATCGAAACGC TGAAACTGCC CAGCTTCGAA
4051 AGCCCTACTC CGCCCAAACT GACCGCCCCC GGTGGCTATA TCGTCGACAT
4101 TCCGAAAGGC AATTTGAAAA CCGAAATCGA AAAGCTGGCC AAACAGCCCG
4151 AGTATGCTTA TCTGAAACAG CTCCAAGTAG CGAAAAACGT CAACTGGAAC
4201 CAGGTGCAAC TGGCTTACGA TAAATGGGAC TATAAGCAGG AAGGCTTAAC
4251 CAGAGCCGGT GCAGCGATTG TTACCAATAT CGTAACCGCA CTGACTTATG
4301 GATACGGCGC AACCAGCAGC GGCGGTGTAG CCGCTTCAGG AAGTAGTACA
4351 GCCGACGCTG CCGGAACAGC CGCCACAACG ACAGCAGCAG CTACTACCGT
4401 TTCTACAGCG ACTGCCATGC AAACCGCTGC TTTAGCCTCC TTGTATAGCC
4451 AAGCAGCTGT ATCCATCATC AATAATAAAG GTGATGTCGG CAAAGCGTTG
4501 AAAGATCTCG GCACCACTGA TACGGTCAAG CAGATTGTCA CTTCTGCCCT
4551 GACGGCGGGT GCATTAAATC AGATGGGCGC AGATATTGCC CAATTGAACA
4601 GCAAGGTAAG AACCGAACTG TTCAGCAGTA CGGGCAATCA AACTATTGCC
4651 AACCTTGGAG GCAGACTGGC TACCAATCTC AGTAATGCAG GTATCTCAGC
4701 TGGTATCAAT ACCGCCGTCA ACGGCGGCAG CCTGAAAGAC AACTTAGGCA
4751 ATGCCGCATT AGGAGCATTG GTTAATAGCT TCCAAGGAGA AGCCGCCAGC
4801 AAAATCAAAA CAACCTTCAG GCACGATTAT GTTGCCAAAC AGTTCGCCCA

4851 CGCTTTGGCT GGGTGTGTTA GCGGATTGGT ACAAGGAAAA TGTAAGACG
 4901 GGGCAATTGG CGCAGCAGTT GGGGAAATCG TAGCCGACTC CATGCTTGGC
 4951 GGCAGAAACC CTGCTACACT CAGCGATGCG GAAAAGCATA AGGTTATCAG
 5001 TTACTCGAAG ATTATTGCCG GCAGCGTGGC GGCATCAAC GCGGGCGATG
 5051 TGAATCTGC GCGGAATGCG GCTGAGGTGG CGGTAGTGAA TAATGCTTTG
 5101 AATTTTGACA GTACCCCTAC CAATGCGAAA AAGCATCAAC CGCAGAAGCC
 5151 CGACAAAACC GCACTGGAAA AAATTATCCA AGGTATTATG CCTGCACATG
 5201 CAGCAGGTGC GATGACTAAT CCGCAGGATA AGGATGCTGC CATTTGGATA
 5251 AGCAATATCC GTAATGGCAT CACAGGCCCG ATTGTGATTA CCAGCTATGG
 5301 GGTTTATGCT GCAGGTTGGA CAGCTCCGCT GATCGGTACA GCGGGTAAAT
 5351 TAGCTATCAG CACCTGCATG GCTAATCCTT CTGGTTGTAC TGTCTATGGT
 5401 ACTCAGGCTG CCGAAGCGGG CGCGGGGATC GCCACGGGTG CGGTAACGGT
 5451 AGGCAACGCT TGGGAAGCGC CTGTGGGGGC GTTGTGCGAA GCGAAGCGCG
 5501 CCAAGCAGGC TATACCAACC CAGACAGTTA AAGAAGTTGA TGGCTTACTA
 5551 CAAGATCAA AAAATATAGG TGCTGTAAAT ACACGAATTA ATATAGCGAA
 5601 TAGTACTACT CGATATACAC CAATGAGACA AACGGGACAA CCGGTATCTG
 5651 CTGGCTTTGA GCATGTTCTT GAGGGGCACT TCCATAGGCC TATTGCGAAT
 5701 AACCCTTCAG TTTTACCAT CTCCCCAAAT GAATTGAAGG TTATACTTCA
 5751 AAGTAATAAA GTAGTTTCTT CTCCCGTATC GATGACTCCT GATGGCCAAT
 5801 ATATGCGGAC TGTCGATGTA GGAAAAGTTA TTGGTACTAC TTCTATTAAA
 5851 GAAGTGGAC AACCACAAC TACAATTAAA GTATTTCAG ATAAGTCAGG
 5901 AAATTTGATT ACTACATACC CAGTAAAGG AACTAA

This corresponds to the amino acid sequence <SEQ ID 60; ORF114-1>:

1 MNKGLHRIIF SKKHSTMVAV AETANSQKGK KQAGSSVSVS LKTSGLDCGK
 51 LKTTLKTLCV SLVSLSMVLP AHAQITDKS APKNQQVVIL KNTNGAPLVN
 101 IQTPNGRGLS HNRYTQFDVD NKGAVLNDR NNNPFVVKGS AQLILNEVRG
 151 TASKLNGIVT VGGQKADVII ANPNGITVNG GGFKNVGRGI LTTGAPQIGK
 201 DGALTGFVDR QGTLTVGAAG WNDKGGADYT GVLARVALQ GKLQGNLAV
 251 STGPQKVDYA SGEISAGTAA GTKPTIALDT AALGGMYADS ITLIANEKGV
 301 GVKNAGTLEA AKQLIVTSSG RIENSRIAT TADGTEASPT YLSIETTEKG
 351 AAGTFISNGG RIESKGLLVI ETGEDISLRN GAVVQNGSR PATTVLNAGH
 401 NLVIESKTNV NNAKGPATLS ADGRTVIKEA SIQTGTVVYS SSKGNAELGN
 451 NTRITGADVT VLSNGTISSS AVIDAKDTAH IEAGKPLSLE ASTVTSDIRL
 501 NGGSLKGGKQ LALLADNIT AKTTNLNTPG NLYVHTGKDL NLNVDRDLA
 551 ASIHLSKDNA AHITGTSKTL TASKDMGVEA GSLNVTNTNL RTNSGNLHIQ
 601 AAKGNIQLRN TKLNAAKALE TTALQGNIVS DGLHAVSADG HVSLLANGNA
 651 DFTGHNTLTA KADVNAAGSVG KGRLLKADNTN ITSSSGDITL VAGNGIQLGD
 701 GKQRNSINGK HISIKNGGAD ADLKNLNVHA KSGALNHSD RALSIENTKL
 751 ESTHNTLNA QHERVTNLQV DAYAHRHLSI TGSQIWQNDK LPSANKLVAN
 801 GVLALNARYS QIADNTTLRA GAINLTAGTA LVKRGNIWS TVSTKTLEDN
 851 AELKPLAGRL NIEAGSGTLT IEPANRISAH TDLISIKTGGK LLSAKGNA
 901 GAPSAQVSSL EAKGNIRLVT GETDLRGSKI TAGKNLVVAT TKGKLNIEAV
 951 NNSFSNYFPT QKAAELNQKS KELEQQIAQL KRSSPKSKLI PTLQEERDRL
 1001 AFYIQAINKE VKGKKPKGKE YLQAKLSAQN IDLISAQIE ISGSDITASK
 1051 KLNLAAGVLP PKAADSEAAA ILIDGITDOY EIGKPTYKSH YDKAALNKPS
 1101 RLTRGTGVSI HAAALDDAR IIGASEIKA PSGSIDIKAH SDIVLEAGQN
 1151 DAYTFKTKG KSGKIIRKTK FTSTRDHLIM PAPVELTANG ITLQAGGNIE
 1201 ANTTFRNAPA GKVTLVAGEE LQLLAEEGIH KHELDVQKSR RFIGIKVGKS
 1251 NYSKNEINET KLPVRVVAQT AATRSWDTV LEGTEFKTTL AGADIQAGVG
 1301 EKARADAKII LKGIVNRIQS EEKLETNSTV WQKQAGRST IETLKLPSFE
 1351 SPTPPKLTAP GGYIVDIPKG NLKTEIEKLA KQPEYAYLQ LQVAKNVNWN
 1401 QVQLAYDKWD YKQEGLTRAG AAIVTIIVTA LTYGYGATAA GGVAAAGSST
 1451 AAAAGTAATT TAAATTVSTA TAMQTAALAS LYSQAAVSII NNKGDVKGAL
 1501 KDLGTSDTVK QIVTSALTAG ALNQMADIA QLNKSVRTEL FSSTGNQITIA
 1551 NLGRLATNL SNAGISAGIN TAVNGGSLKD NLGNAALGAL VNSFQGEAAS
 1601 KIKTTFSDDY VAKQFAHALA GCVSGLVQVK CKDGAIGAAV GEIVADSMGL
 1651 GRNPATLSDA EKHKVISYSK IAGSVAALN GGDVNTAANA AEVAVVNNAL
 1701 NFDSTPTNAK KHQPKPKDKT ALEKIIQIGM PAHAAGAMTN PQDKDAIWI
 1751 SNIRNGITGP IVITSYGVYA AGWTAPLIGT AGKLAISTCM ANPSGCTVMV
 1801 TQAAEAGAGI ATGAVTVGNA WEAPVGALSK AKAQKAIPT QTVKELDGLL
 1851 QESKNIGAVN TRINIANSTT RYTPMRQTGQ PVSAGFEHVL EGFHFRPIAN
 1901 NRSVFTISPN ELKVILQSNK VVSSPVSMTP DGQYMRITVDV GKVIGTTSIK
 1951 EGGQPTTTIK VFTDKSGNLI TTPVVKGN*

Computer analysis of this amino acid sequence predicts a transmembrane region and also gives the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF114 shows 91.9% identity over a 284aa overlap with an ORF (ORF114a) from strain A of *N.*

meningitidis:

5	orf114.pep	AVAETANSQKGKQAGSSVSLSKTSGLDLCGKLKTTTLKTLVC	10	20	30	40
	orf114a	MNKGHLHRIIFSCKHSTMVAEAVETANSQKGKQAGSSVSLSKTSGLDLCGKLKTTTLKTLVC	10	20	30	40
10	orf114.pep	SLVSLSMVLPAAHAQITTDKSAPKNQVVLKTNLTGAPLVNIQTPNRGRLSHNRXYAFDVD	50	60	70	80
	orf114a	SLVSLSMXXXXXXQITTDKSAPKNQVVLKTNLTGAPLVNIQTPNRGRLSHNRXYQFDVD	70	80	90	100
15	orf114.pep	NKGAVLNNDNRNNNPFVVKGSQQLILNEVRGTASKLNGIVTVGGQKADVIIANPNGITVNG	110	120	130	140
	orf114a	NKGAVLNNDNRNNNPFVVKGSQQLILNEVRGTASKLNGIVTVGGQKADVIIANPNGITVNG	130	140	150	160
20	orf114.pep	GGFKNVGRGILTTGAPQIGKDGALTGFDDVVKAHWTVXAAGWNDKGGAXYTGVLARAVALQ	170	180	190	200
	orf114a	GGFKNVGRGILTTGAPQIGKDGALTGFDDVVKAHWTVXAAGWNDKGGADYTGVLARAVALQ	190	200	210	220
25	orf114.pep	GKXXGKXLAIVSTGPKVDYASGEISAGTAAGTKPTIALDTAALGMYADSLTIANEKGKV	230	240	250	260
	orf114a	GKLGKXLAIVSTGPKVDYASGEISAGTAAGTKPTIALDTAALGMYADSLTIAEKGKV	250	260	270	280
30	orf114.pep	GVX	290	300		
	orf114a	GVKNAGTLEAAKQLIVTSSGRIENSGRIATTADGTEASPTYLXIETTEKGAXGTFISNGG	310	320	330	340
35			350	360		

The complete length ORF114a nucleotide sequence <SEQ ID 61> is:

40	1	ATGAATAAAG	GTTTACATCG	CATTATCTTT	AGTAAAAAGC	ACAGCACCAT
	51	GGTTGCAGTA	GCCGAAACTG	CCAACAGCCA	GGGCAAAGGT	AAACAGGCAG
	101	GCAGTTCGGT	TTCTGTTTCA	CTGAAAACTT	CAGGCGACCT	TTGCGGCAAA
	151	CTCAAAACCA	CCCTTAAAC	CTTGGTCTGC	TCTTTGGTTT	CCCTGAGTAT
	201	GGNATTNCNN	NNCNTNCCC	AAATTACCAC	CGACAAATCA	GCACCTAAAA
45	251	ACCANCAGGT	CGTTATCCTT	AAAACCAACA	CTGGTGCCCC	CTTGGTGAAT
	301	ATCCAAACTC	CGAATGGACG	CGGATTGAGC	CACAACCGCT	ATACGCAGTT
	351	TGATGTTGAC	AACAAAGGGG	CAGTGTAAAT	CAACGACCGT	AACAATAATC
	401	CGTTTCTGGT	CAAAGGCAGT	GCGCAATTGA	TTTTGAACGA	GGTACGCGGT
	451	ACGGCTAGCA	AACTCAACGG	CATCGTTACC	GTAGGCGGTC	AAAAGGCCGA
50	501	CGTGATTATT	GCCAACCCCA	ACGGCATTAC	CGTTAATGGC	GGCGGCTTTA
	551	AAAATGTCCG	TCGGGGCATC	TTAACTATCG	GTGCGCCCCA	AATCGGCAAA
	601	GACGGTGCAC	TGACAGGATT	TGATGTGCGT	CAAGGCACAT	TGACCGTAGG
	651	AGCAGCAGGT	TGGAATGATA	AAGGCGGAGC	CGACTACACC	GGGGTACTTG
	701	CTCGTGCACT	TGCTTTGCAG	GGGAAATTAC	AGGGTAAAAA	CCTGGCGGTT
55	751	TCTACCGGTC	CTCAGAAAGT	AGATTACGCC	AGCGGCGAAA	TCAGTGCAGG
	801	TACGGCAGCG	GGTACGAAAC	CGACTATTGC	CCTTGATACT	GCCGCACTGG
	851	GCGGTATGTA	CGCCGACAGC	ATCACACTGA	TTGCCANTGA	AAAAGGCGTA
	901	GGCGTCAAAA	ATGCCGGCAC	ACTCGAAGCG	GCCAAGCAAT	TGATTGTGAC
	951	TTCGTGAGGC	CGCATGTAAA	ACAGCGGCCG	CATCGCCACC	ACTGCCGACG
60	1001	GCACCGAAGC	TTCACTGACT	TATCTNNCNA	TGCAAAACCAC	CGAAAAAGGA
	1051	GCNNCAGGCA	CATTTATCTC	CAATGGTGGT	CGGATCGAGA	GCAAAGGCTT
	1101	ATTGGTTATT	GAGACGGGAG	AAGATATCAN	CTTGCGTAAC	GGAGCCGTGG
	1151	TGCAGATAAA	CGGCAGTCGC	CCAGCTACCA	CGGTATTAAA	TGCTGGTCAT
	1201	AATTTGGTGA	TTGAGAGTAA	AACTAATGTG	AACAATGCCA	AAGGCTCGNC

5

10

15

20

25

30

35

40

45

50

55

60

65

```

1251 TAATCTGTCG GCCGGCGGTC GTACTACGAT CAATGATGCT ACTATTCAAG
1301 CGGGCAGTTC CGTGACAGC TCCACCAAAG GCGATACTGA NTTGGGTGAA
1351 AATACCCGTA TTATTGCTGA AAACGTAACC GTATTATCTA ACGGTAGTAT
1401 TGGCAGTGCT GCTGTAATTG AGGCTAAAGA CACTGCACAC ATTGAATCGG
1451 GCAAACCGCT TTCTTTAGAA ACCTCGACCG TTGCCTCCAA CATCCGTTTG
1501 AACAAACGGTA ACATTAAAGG CGGAAAGCAG CTTGCTTTAC TGGCAGACGA
1551 TAACATTACT GCCAAACTA CCAATCTGAA TACTCCCGGC AATCTGTATG
1601 TTCATACAGG TAAAGATCTG AATTGGAATG TTGATAAAGA TTTGTCTGCC
1651 GCCAGCATCC ATTTGAAATC GGATAACGCT GCCCATATTA CCGGCACCAG
1701 TAAACCCCTC ACTGCCTCAA AAGACATGGG TGTGGAGGCA GGCTTGCTGA
1751 ATGTTACCAA TACCAATCTG CGTACCAACT CGGGAATCT GCACATTGAG
1801 GCAGCCAAAG GCAATATTCA GCTTCGCAAT ACCAAGCTGA ACGCAGCCAA
1851 GGCTCTCGAA ACCACGCAT TGCAGGGCAA TATCGTTTCA GACGGCCTTC
1901 ATGCTGTTTC TGCAGACGGT CATGTATCCT TATTGGCCAA CGGTAATGCC
1951 GACTTTACCG GTCACAATAC CCTGACAGCC AAGGCGGATG TCNATGCAGG
2001 ATCGGTTGGT AAAGGCCGTC TGAAAGCAGA CAATACCAAT ATCACTTCAT
2051 CTTACAGGAG TATTACGTTG GTTGCCGNNN NCGGTATTCA GCTTGGTGAC
2101 GGAAACAAC GCAATTCAAT CAACGGAAAA CACATCAGCA TCAAAAACAA
2151 CGGTGGTAAT GCCGACTTAA AAAACCTTAA CGTCCATGCC AAAAGCGGGG
2201 CATTGAACAT TCATTCCGAC CGGGCATTGA GCATAGAAAA TACNAAGCTG
2251 GAGTCTACCC ATAATACGCA TCTTAATGCA CAACACGAGC GGGTAACGCT
2301 CAACCAAGTA GATGCCTACG CACACCGTCA TCTAAGCATT ANCGGCAGCC
2351 AGATTGGGCA AAACGACAAA CTGCCTTCTG CCAACAAGCT GGTGGCTAAC
2401 GGTGTATTGG CANTCAATGC GCGCTATTCC CAAATTGCCG ACAACACCAC
2451 GCTGAGAGCG GGTGCAATCA ACCTTACTGC CGGTACCGCC CTAGTCAAGC
2501 GCGGCAACAT CAATTGGAGT ACCGTTTCGA CCAAGACTTT GGAAGATAAT
2551 GCCGAATTAA AACCATTGGC CGGACGGCTG AATATTGAAG CAGGTAGCGG
2601 CACATTAAAC ATCGAACCTG CCAACCGCAT CAGTGGCAT ACCGACCTGA
2651 GCATCAAAAC AGGCGGAAAA TTGCTGTTGT CTGCAAAAGG AGGAAATGCA
2701 GGTGCCNTA GTGCTCAAGT TTCCTCATTG GAAGCAAAAG GCAATATCCG
2751 TCTGGTTACA GGAGNAACAG ATTTAAGAGG TTCTAAAATT ACAGCCGGTA
2801 AAAACTTGGT TGTGCGCCAC ACCAAAGGCA AGTTGAATAT CGAAGCCGTA
2851 AACAACCTCAT TCAGCAATTA TTTTCNTACA CAAAAGNGN NNGNNCTCAA
2901 CCAAAATCC AAAGAATTGG AACAGCAGAT TGCGCAGTTG AAAAAAGCT
2951 GCNTAAAAG CAAGCTGATT CCAACCTGTC AAGAAGAAGC CGACCGTCTC
3001 GCTTTCTATA TTCAAGCCAT CAACAAGGAA GTTAAAGGTA AAAAACCCAA
3051 AGGCAAGAA TACCTGCAAG CCAAGCTTTC TGCACAAAT ATTGACTTGA
3101 TTTCCGCACA AGGCATCGAA ATCAGCGGTT CCGATATTAC CGCTTCCAAA
3151 AAACCTGAAC TTCACGCGCG AGGCGTATTG CCAAAGGCAG CAGATTGAGA
3201 GGCGGCTGCT ATTCTGATTG ACGGCATAAC CGACCAATAT GAAATTGGCA
3251 AGCCCCACTA CAAGAGTCAC TACGACAAAG CTGCTCTGAA TAGCCTTCA
3301 CGTTTGACCG GACGTACGGG GGTAAGTATT CATGCAGCTG CGGCACTCGA
3351 TGATGCACGT ATTATTATCG GTGCATCCGA AATCAAAGCT CCTCAGGCA
3401 GCATAGACAT CAAAGCCCAT AGTGATATTG TACTGGAGGC TGGACAAAAC
3451 GATGCCTATA CCTTCTTANA AACCAAAGGT AAAAGCGGCA NAATNATCAG
3501 AAAAACNAAG TTACCAGCA CCNCGANCA CTTGATTATG CCAGCCCCNG
3551 TCGAGCTGAC CGCCAACGGT ATCAGCTTTC AGGCAGGCGG CAACATCGAA
3601 GCTAATACCA CCGCTTCAA TGCCCTGCA GGTAAAGTTA CCCTGGTTGC
3651 GGGTGAANAG NTGCAACTGC TGGCAGAAGA AGGCATCCAC AAGCAGAGT
3701 TGGATGTCCA AAAAAGCCGC CGCTTTATCG GCATCAAGT AGGTNAGAGC
3751 AATTACAGTA AAACGAACT GAACGAAACC AAATTGCCTG TCCGCGTCTG
3801 CGCCCAAANT GCAGCCACCC GTTCAGGCTG GGATACCGTG CTCGAAGGTA
3851 CCGAATTCAA AACCACGCTG GCCGGTGCCG ACATTAGGC AGGTGTANGC
3901 GAAAAAGCCC GTGTCGATGC GAAAATTATC CTCAAAGGCA TTGTGAACCG
3951 TATCCAGTCG GAAGAAAAAT TAGAAACCAA CTCAACCGTA TGGCAGAAAC
4001 AGGCCGGACG CGGCAGCACT ATCGAAACGC TAAACTGCC CAGCTTCGAA
4051 AGCCCTACTC CGCCCAAATT GTCCGCACCC GCGGNTATA TCGTCGACAT
4101 TCCGAAAGGC AATCTGAAAA CCGAAATCGA AAAGCTGTCC AAACAGCCCG
4151 AGTATGCCTA TCTGAAACAG CTCCAAGTAG CGAAAAACAT CAACTGGAAT
4201 CAGGTGCAGC TTGCTTACGA CAGATGGGAC TACAAACAGG AGGGCTTAAC
4251 CGAAGCAGGT GCGGCGATTA TCGCACTGGC CGTTACCGTG GTCACCTCAG
4301 GCGCAGGAAC CGGAGCCGTA TTGGGATTAA ACGGTGCGNC CGCCGCCGCA
4351 ACCGATGCAG CATTGCGCTC TTTGGCCAGC CAGGCTTCCG TATCGTTTAT
4401 CAACAACAAA GCGGATGTCG GCAAAACCCT GAAAGAGCTG GGCAGAAGCA
4451 GCACGGTGAA AAATCTGGTG GTTGCCGCGC CTACCGCAGG CGTAGCCGAC
4501 AAAATCGGCG CTTCCGCACT GANCAATGTC AGCGATAAGC AGTGGATCAA
4551 CAACCTGACC GTCAACCTAG CCAATGNCAG CAGTGCCGC ACTGattaa

```

This encodes a protein having amino acid sequence <SEQ ID 62>:

1 MNKGLHRIIF SKKHSTMVAV AETANSQGKG KQAGSSVSVS LKTSGLCGK

51 LKTTTLKTLVC SLVSLSMXXX XXXQITTDKS APKNXQVVIL KTNTGAPLVN
 101 IQTPNGRGLS HNRYTQFDVD NKGAVLNDR NNNPFLVKGS AQLILNEVRG
 151 TASKLNGIVT VGGQKADVII ANPNGITVNG GGFKNVGRGI LTIGAPQIGK
 201 DGALTGFQDVR QGTLTVGAAG WNDKGGADYT GVLARAVALQ GKLGKKNLAV
 251 STGPQKVDYA SGEISAGTAA GTKPTIALDT AALGGMYADS ITLIAHEKGV
 301 GVKNAGTLEA AKQLIVTSSG RIENSGRIAT TADGTEASPT YLXIETTEKG
 351 AXGTFISNGG RIESKGLLVI ETGEDIXLRN GAVVQNNNGSR PATTVLNAGH
 401 NLVIESKTNV NNAKGSXNLS AGGRTTINDA TIQAGSSVYS STKGDTXLGE
 451 NTRIIAENVV VLSNGSIGSA AVIEAKDTAH IESGKPLSLE TSTVASNIRL
 501 NNGNIKGGKQ LALLADDNIT AKTTNLNTPG NLYVHTGKDL NLNVDKDLSA
 551 ASIHLSKSDNA AHITGTSKTL TASKDMGVEA GLLNVNTNLT RTNSGNLHIQ
 601 AAKGNIQLRN TKLNAKALE TTALQGNIVS DGLHAVSADG HVSLLANGNA
 651 DFTGHTLTA KADVXAGSVG KGRKADNTN ITSSSGDITL VAXXGIGLGD
 701 GKQNSINGK HISIKNNGGN ADLKNLVNHA KSGALNIHSD RALSIENTKL
 751 ESTHNTLNA QHERVTINQV DAYAHRHLSI XGSQWQNDK LPSANKLVAN
 801 GVLAXNARYS QIADNTTLRA GAINLTAGTA LVKRGNNWS TVSTKTLEDN
 851 AELKPLAGRL NIEAGSGTTL IEPANRISAH TDLSIKTGGK LLLSAGGNA
 901 GAXSAQVSSL EAKGNIRLVT GXTDLRGSKI TAGKNLVVAT TKGKLNIEAV
 951 NNSFSNYFXT QKXXKLNQKS KELEQQIAQL KSSXKSKLI PTLQEEEDRL
 1001 AFYIQAENKE VKGKPKGKE YLQAKLSAQN IDLISAQGIE ISGSDITASK
 1051 KLNLAAGVL PKAADSEAAA ILIDGITDQY EIGKPTYKSH YDKAALNKPS
 1101 RLTGRTGVSI HAAALDDAR IIIGASEIKA PSGSIDIKAH SDIVLEAGQN
 1151 DAYTFLXTKG KSGXXIRKTK FTSTXHLIM PAPVELTANG ITLQAGGNIE
 1201 ANTTRENAPE GKVTILVAGX KQLLAEEGHI KHELDVQKSR RFIGIKVXKS
 1251 NYSKNEINET KLPVRVVAQX AATRSQWDTV LEGTEFKTTL AGADIQAGVX
 1301 EKARVDAKII LKGIVNRIQS EEKLETNSTV WQKQAGRGST IETLKLPSFE
 1351 SPTPPKLSAP GGYIVDIPKG NLKTEIEKLS KOPEYAYLKQ LQVAKNINWN
 1401 QVQLAYDRWD YKQEGLTEAG AAILALAVTV VTSGAGTGAV LGLNGAXAAA
 1451 TDAAFASLAS QASVSFINNK GDVGKTLKEL GRSSTVKNLV VAAATAGVAD
 1501 KIGASALXNV SDKQWNNLT VNLANKQCR TD*

ORF114-1 and ORF114a show 89.8% identity in 1564 aa overlap

35 orf114a.pep MNKGLHRIIFSCKHSTMVAVAETANSQKGKQAGSSVSLSKTSGLCGKLTTLKTLVC
 orf114-1 MNKGLHRIIFSCKHSTMVAVAETANSQKGKQAGSSVSLSKTSGLCGKLTTLKTLVC
 40 orf114a.pep SLVSLSMXXXXXXQITTDKSAPKNXQVVILKTNTGAPLVNIQTTPNGRGLSHNRYTQFDVD
 orf114-1 SLVSLSMVLPAAHQITTDKSAPKNQQVVILKTNTGAPLVNIQTTPNGRGLS:RYTQFDVD
 45 orf114a.pep NKGAVLNDRNNNNPFLVKGS AQLILNEVRGTASKLNGIVTVGGQKADVIIANPNGITVNG
 orf114-1 NKGAVLNDRNNNNPFLVKGS AQLILNEVRGTASKLNGIVTVGGQKADVIIANPNGITVNG
 50 orf114a.pep GGFKNVGRGILTTGAPQIGKDGALTGFQDVRQGTTLTVGAAGWNDKGGADYTGVLARAVALQ
 orf114-1 GGFKNVGRGILTTGAPQIGKDGALTGFQDVRQGTTLTVGAAGWNDKGGADYTGVLARAVALQ
 55 orf114a.pep GKLGKKNLAVSTGPQKVDYASGEISAGTAAGTKPTIALDTAALGGMYADSITLIAHEKGV
 orf114-1 GKLGKKNLAVSTGPQKVDYASGEISAGTAAGTKPTIALDTAALGGMYADSITLIAHEKGV
 60 orf114a.pep GVKNAGTLEAAKQLIVTSSGRIENSGRIATTADGTEASPTYLXIETTEKGAXGTFISNGG
 orf114-1 GVKNAGTLEAAKQLIVTSSGRIENSGRIATTADGTEASPTYLSIETTEKGAAGTFISNGG
 65 orf114a.pep RIESKGLLVIETGEDIXLRNGAVVQNNNGSRPATTVLNAGHNLVIESKTNVNNNAKGSXNLS
 orf114-1 RIESKGLLVIETGEDISLRNGAVVQNNNGSRPATTVLNAGHNLVIESKTNVNNNAKGPATLS
 70 orf114a.pep AGGRTTINDATIQAGSSVYSSTKGDTXLGENTRIIAENVTVLSNGSIGSAAVIEAKDTAH
 orf114-1 ADGRTVIKEASIQGTTVYSSSKGNAELGNTRITGADVTVLSNGTSSSAVIDAKDTAH
 75 orf114a.pep IESGKPLSLETSTVASNIRLNNNGNIKGGKQLALLADDNITAKTTNLNTPGNLYVHTGKDL
 orf114-1 IEAGKPLSLEASTVTSIRLNGGSIKGGKQLALLADDNITAKTTNLNTPGNLYVHTGKDL
 80 orf114a.pep NLNVDKDLSAASIHLSKSDNAAHITGTSKTLTASKDMGVEAGLLNVNTNLTNTNSGNLHIQ
 orf114-1 NLNVDKDLSAASIHLSKSDNAAHITGTSKTLTASKDMGVEAGLLNVNTNLTNTNSGNLHIQ

orf114-1 NLNVDKDLSAASIHLSKSDNAAHITGTSKTLTASKDMGVEAGSLNVTNTNLRNNSGNLHIQ
orf114a.pep AAKGNIQLRNTKLNAAKALETTALQGNIVSDGLHAVSADGHVSLLANGNADFTGHNTLTA
5 orf114-1 AAKGNIQLRNTKLNAAKALETTALQGNIVSDGLHAVSADGHVSLLANGNADFTGHNTLTA
orf114a.pep KADVXAGSVGKGRLLKADNTNITSSSGDITLVAXXGIQLGDGKQRNSINGKHISIKNNGGN
10 orf114-1 KADVXAGSVGKGRLLKADNTNITSSSGDITLVAXXGIQLGDGKQRNSINGKHISIKNNGGN
orf114a.pep ADLKNLNVHAKSGALNIHSDRALSIENTKLESTHNTHLNAQHERVTLNQVDAYAHRHLSI
orf114-1 ADLKNLNVHAKSGALNIHSDRALSIENTKLESTHNTHLNAQHERVTLNQVDAYAHRHLSI
15 orf114a.pep XGSQIWQNDKLP SANKLVANGVLAXNARYSQIADNTTLRAGAINLTAGTALVKRGNINWS
orf114-1 XGSQIWQNDKLP SANKLVANGVLAXNARYSQIADNTTLRAGAINLTAGTALVKRGNINWS
20 orf114a.pep TVSTKTLEDNAELKPLAGRLNIEAGSGTLTIEPANRISAHTDLSIKTGGKLLLSAKGGNA
orf114-1 TVSTKTLEDNAELKPLAGRLNIEAGSGTLTIEPANRISAHTDLSIKTGGKLLLSAKGGNA
orf114a.pep GAXSAQVSSLEAKGNIRLVGTGXTDLRSGKITAGKNLVVATTKGKLNIEAVNNSFSNYFXT
25 orf114-1 GAXSAQVSSLEAKGNIRLVGTGXTDLRSGKITAGKNLVVATTKGKLNIEAVNNSFSNYFXT
orf114a.pep QKXXXLNQKSKELEQQIAQLKKSSXKSKLIPTLQEERDLAFYIQAINKEVKGKKPKGKE
orf114-1 QKXXXLNQKSKELEQQIAQLKKSSXKSKLIPTLQEERDLAFYIQAINKEVKGKKPKGKE
30 orf114a.pep YLQAKLSAQNIIDLISAQGIEISGSDITASKKLNLAAGVLPKAADSEAAAILIDGITDQY
orf114-1 YLQAKLSAQNIIDLISAQGIEISGSDITASKKLNLAAGVLPKAADSEAAAILIDGITDQY
35 orf114a.pep EIGKPTYKSHYDKAALNKP SRTLGRGTGVSIAHAAALDDARIIGASEIKAPSGSIDIKAH
orf114-1 EIGKPTYKSHYDKAALNKP SRTLGRGTGVSIAHAAALDDARIIGASEIKAPSGSIDIKAH
40 orf114a.pep SDIVLEAGQNDAYTFLXTKGKSGXIRKTKFTSTXXHLIMPAPVELTANGITLQAGGNIE
orf114-1 SDIVLEAGQNDAYTFLXTKGKSGXIRKTKFTSTXXHLIMPAPVELTANGITLQAGGNIE
orf114a.pep ANTTRFNAPAGKVTILVAGEEXQLLAEEGIHKHELDVQKSRRFIGIKVGXSNYSKNELNET
45 orf114-1 ANTTRFNAPAGKVTILVAGEEXQLLAEEGIHKHELDVQKSRRFIGIKVGXSNYSKNELNET
orf114a.pep KLPVRVVAQXAATRSQWDTVLEGTETFTTLAGADIQAGVXEKARVDAKIIILKGIVNRIQS
50 orf114-1 KLPVRVVAQXAATRSQWDTVLEGTETFTTLAGADIQAGVXEKARVDAKIIILKGIVNRIQS
orf114a.pep EEKLETNSTVWQKQAGRGSTIETLKLPSFESPTPPKLSAPGGYIVDIPKGNLKEIEKLS
orf114-1 EEKLETNSTVWQKQAGRGSTIETLKLPSFESPTPPKLSAPGGYIVDIPKGNLKEIEKLS
55 orf114a.pep KQPEYAYLKQLQVAKNINWNQVQLAYDRWDYKQEGLTEAGAAIIALAVTVVTSAGGTGAV
orf114-1 KQPEYAYLKQLQVAKNINWNQVQLAYDRWDYKQEGLTEAGAAIIALAVTVVTSAGGTGAV
60 orf114a.pep LGLNGA-----XAAATD-----AAFASLASQASVSFINNKGVDGKTL 1477
orf114-1 LGLNGA-----XAAATD-----AAFASLASQASVSFINNKGVDGKTL 1477
GGVAASGSSTAAAAGTAATTTAAATTVSTATAMQTAALASLYSQAASVIINNKGVDGKAL 1500
orf114a.pep KELGRSSTVKNLVVAAATAGVADKIGA-----SALXNVSDKQWINNL-----TVNL 1523
65 orf114-1 KELGRSSTVKNLVVAAATAGVADKIGA-----SALXNVSDKQWINNL-----TVNL 1523
KDLGTSDTVQIVTSALTAGALNQMGADIAQLNSKVRTELFSSSTGNQTIANLGGRLATNL 1560
orf114a.pep ANXGQCRTDX
orf114-1 ANXGQCRTDX
70 orf114a.pep SNAGISAGINTAVN...

Homology with *pspA* putative secreted protein of *N.meningitidis* (accession number AF030941)

ORF114 and *pspA* protein show 36% aa identity in 302aa overlap:

```

Orf114: 1  AVAETANSQKKGKQAGSSVSLSL-----KTSGDXXXXXXXXXXXXXXXXXXXXXPAHAQ 56
          AVAE + GK Q + SV + S PA A
5  pspA: 19  AVAENVHRDGKSMQDSEAAASVRVTGAASVSSARAAGFGRMAAFVMLALGVAAFSPAPAS 78

Orf114: 57  -ITTDKSAPKNQOVVILKNTNTGAPLVNIQTPNGRGLSHNRXYAFDVDNKGAVLNDRNN- 114
          I DKSAPKNQ VIL+T G P VNIQTP+ +G+S NR FDVD KG +LNN R+N
10 pspA: 79  GIIADKSAPKNQQAAILQTANGLPQVNIQTSSQGVSVNRFKQFDVDEKGVILNNSRSNT 138

Orf114: 115  -----NPFVVKGSQAQLILNEV-RGTASKLNGIVTVGGQKADVIIANPNGITVNGG 163
          NP + +G A++I+N++ S LNG + VGG++A+V++ANP+GI VNGG
15 pspA: 139 QTQLGGWIOGNPHLARGEARVIVNQIDSSNPSLLNGYIEVGGKRAEVVNPSPGIRVNGG 198

Orf114: 164  GFKNVGRGILTGTGAPQIGKDGALTGFDDVVKAHWTVXAAGWNDKGGAXYTGVLARAVALQG 223
          G N LT+G P + +G LTGFDV + G D A YT +L+RA +
20 pspA: 199 GLINAASVTLTSGVPVL-NNGNLTGFDVSSGKVVIGGKGL-DTSDADYTRILSRAAEINA 256

Orf114: 224  KXXGKXLAVSTGPKQVDYASGEISAGTAAGTK-----PTIALDTAALGGMYADSITLIANE 279
          GK + V +G K+D+ +A + PT+A+DTA LGGMYAD ITLI+ +
25 pspA: 257 GVGKDVVKVSGKNKLDGSLAKTASAPSSSDSVTPTVAIDTATLGGMYADKITLISTD 316

Orf114: 280  KG 281
          G
25 pspA: 317 NG 318

```

ORF114a is also homologous to *pspA*:

```

gi|2623258 (AF030941) putative secreted protein [Neisseria meningitidis] Length
= 2273
30 Score = 261 bits (659), Expect = 3e-68
   Identities = 203/663 (30%), Positives = 314/663 (46%), Gaps = 76/663 (11%)

Query: 1  MNKGLHRIIFSKKHSTMVAVAETANSQKKGKQAGSSVSLSL-----TSGDXXXXXXXXXX 55
          MNK +++IF+KK S M+AVAE + GK Q + SV + +S
35 Sbjet: 1  MNKRCYKVIFFNKKRSCMMAVAENVHRDGKSMQDSEAAASVRVTGA/VSSARAAGFGRMAA 60

Query: 56  XXXXXXXXXXXXXXXXXXXXXQITTDKSAPKNQOVVILKNTNTGAPLVNIQTPNGRGLSHNRYT 115
          I DKSAPKN Q VIL+T G P VNIQTP+ +G+S NR+
40 Sbjet: 61  FSVMLALGVAAFSPAPASGIIADKSAPKNQQAAILQTANGLPQVNIQTSSQGVSVNRFK 120

Query: 116  QFDVDNKGAVLNDRNN-----NPFVVKGSQAQLILNEV-RGTASKLNGIVTVGG 163
          QFDVD KG +LNN R+N NP L +G A++I+N++ S LNG + VGG
45 Sbjet: 121  QFDVDEKGVILNNSRSNTQTQLGGWIOGNPHLARGEARVIVNQIDSSNPSLLNGYIEVGG 180

Query: 164  QKADVIIANPNGITVNGGGFKNVGRGILTGTGAPQIGKDGALTGFDDVROGTLTVGAAGWND 223
          ++A+V++ANP+GI VNGG N LT G P + +G LTGFDV G + +G G D
50 Sbjet: 181  KRAEVVNPSPGIRVNGGGLINAASVTLTSGVPVL-NNGNLTGFDVSSGKVVIGGKGL-D 238

Query: 224  KGGADYTGVLARAVALQKLGKLNLA VSTGPKQVDYASGEISAGTAAGTK-----PTIALD 279
          ADYT +L+RA + + GK++ V +G K+D+ +A + PT+A+D
55 Sbjet: 239  TSDADYTRILSRAAEINAGVWGKDVVKVSGKNKLDGSLAKTASAPSSSDSVTPTVAID 298

Query: 280  TAALGGMYADSITLIAXEKGVGKVNAGTLEAAK-QLIVTSSGRIENSGRIATTADGTEAS 338
          TA LGGMYAD ITLI+ + G ++N G + AA + +++ G++ NSG I +A+
60 Sbjet: 299  TATLGGMYADKITLISTDNGAVIRNKGRIFAATGGVTLSDGKLSNSGSI-----DAA 351

Query: 339  PTYLXIETTEKGXGTFFISNGGRIESKGLLVIIETGEDIXLRNGAVVQNNGSRPATTVINA 398
          + +T + + G I S V++ + I + G + GS + +
65 Sbjet: 352  EITISAQTVD-----NRQGFIRSGKSVLKVSDGINNQAGLI-----GSAGLLDIRDT 399

Query: 399  GHNLVIESKTNVNNAGKS-----XNLSAGGRTTINDATIOAGSSVYSSTKGD TXLGENTRI 454
          G +S ++NN G+ ++S ++ ND + A V S + D G+
70 Sbjet: 400  G-----KSSLHINNTDGTIIAGKDVSLQAKSLDNDGILTAARDV-SVSLHDDFAGKRDIE 453

Query: 455  IAENVTVLSNGSIGSAAVIEAKDTAHIESGKPLSLETSTVASNIRLNNGNIKGGKQLALL 514
          +T + G + + +I+A DT + + + + + S R G L+

```

Sbjct: 454 AGRTLTFSTQGR LKNTRIIQAGDTVSLTAAQIDNTVSGKIQSGNRTGLNGKNGITNRGLI 513

Query: 515 ADDNIT-----AKTTNLNTPGNLYVHTGKDLNINVDKDLASAASIHLSKSDNAAHITGTSKT 569
+ IT AK+ N T G +Y G + + D L+ AA

5 Sbjct: 514 NSNGITLLQTEAKSDNAGT-GRIY---GSRVAVEADTLLNREETVNGETKAA-----V 562

Query: 570 LTASKDMGVEAGXXXXXXXXXXXXSGNLHIQAA---KGNIQLRNTKL-NAAKALETALQ 625
+ A + + + A SG+LHI +A +Q NT L N + A+E++

10 Sbjct: 563 IAARERLDIGAREIENREAALLSSSGDLHIGSALNGSRQVQAGANTSLHNRSAIESS---- 619

Query: 626 GNI 628
GNI

Sbjct: 620 GNI 622

15 Score = 37.5 bits (85), Expect = 0.53
Identities = 87/432 (20%), Positives = 159/432 (36%), Gaps = 62/432 (14%)

Query: 239 LQKLOGKNLAVSTGPQKVDYASGEISAGTAAGTKPTIALDTAALGGMYSITLIXEK 298
LQG LQGN+ + G + +G I A A K A + + S T +

20 Sbjct: 1023 LQGDLOGKNIFAAAGSDITN--TGSIGAENALLK-----ASNIESRSETRSNQNE 1072

Query: 299 GVGKVNAGTLEAAKQLIVTSSGRI--ENSGRIATTADGTEASPTYLXIETTEKGAGX-TF 355
V+N G + A L +G+ + + I TA ET + G T

25 Sbjct: 1073 QGSVRNIGRV-AGIYLTGRQNGSVLLDAGNNIVLTAS-----ELTNQSEDGQTV 1120

Query: 356 ISNGGRIESKGLLVITGEDIXLRNGAVVQNGSRPATTVLNAGHNLVIESK-----T 408
++ GG I S + I + V++ + +T+ G NL + +K

30 Sbjct: 1121 LNAGGDIRSDTTGISRNQNTIFDSNYVIRKEQNEVGSTIRTRG-NLSLNAGKDIRIRAA 1179

Query: 409 NVNNAKGSXNLSAGGRITINDATIOAGSS-----VYSSTKGD TXLGENTRIIAENV 460
V + +G L+AG D ++AG + Y+ G + TR +

Sbjct: 1180 EVGSEQGRKLAAAG-----RDIKVEAGKAHTETEDALKYTGSGGGIKQKMRHLKNQNG 1234

35 Query: 461 VLSNGSIGSAAVIEAKDTAHIESGKPLSLETSTVASNIRLNNGNIKGGKQLALLADDNIT 520
+G++ +I +G + + T+ S NN +K + + A+ N

Sbjct: 1235 QAVSGTLDGKEIILVSGRDITVTGSNIIADNHTILS--AKNNIVLKAETRSRSAEMNKK 1292

Query: 521 AKTTNLNTPG-NLYVHTGKDLNINVDKDLASAASIHLSKSDN-----AAHITGTSKTLTA 572
K+ + + G + KD N + +S + S N H T T T+++

40 Sbjct: 1293 EKSGLMGSGGIGFTAGSKKDTQTNRSETVSHTESVVGSLNGNTLISAGKHYTQTGTSYI S 1352

Query: 573 SK-DMGVEAGXXXXXXXXXXXXSGNLHIQAAKG-----NIQLRNTKLNAAKALETALQ 626
+ D+G+ +G + + KG ++ + NT + A A++ G

45 Sbjct: 1353 PQGDVGISSGKISIDAAQNRYSQESKQVYEQKGVTVVAISVPVNTVMGAVDKAVQTVG 1412

Query: 627 NIVSDGLHAVSA 638
+ ++A++A

50 Sbjct: 1413 KSKNSRVNMAAA 1424

Amino acids 1-1423 of ORF114-1 were cloned in the pGex vector and expressed in *E.coli*, as described above. GST-fusion expression was visible using SDS-PAGE, and Figure 5 shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF114-1.

Based on these results, including the homology with the putative secreted protein of *N.meningitidis* and on the presence of a transmembrane domain, it is predicted that this protein from *N.meningitidis*, and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 14

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 63>


```

1  ..CGCTTCATTC ATGATGAAGC AGTCGGCAGC AACATCGGCG GCGGCAAAAT
51  GATTGTTGCA GCCGGGCAGG ATATCAATGT ACGCGGCAnA AGCCTTATTT
101 CTGATAAGGG CATTGTTTTA AAAGCAGGAC ACGACATCGA TATTTCTACT
151 GCCCATAATC GCTATACCGG CAATGAATAC CACGAGAGCA WAAAWTCAGG
5  201 CGTCATGGGT ACTGGCGGAT TGGGCTTTAC TATCGGTAAC CGGAAACTA
251 CCGATGACAC TGATCGTACC AATATTGTsC ATACAGGCAG CATTATAGGC
301 AGCCTGAaTG GAGACACCGT TACAGTTGCA GGAACCGCT ACCGACAAAC
351 CGGCAGTACC GTCTCCAGCC CCGAGGGGCG CAATACCGTC ACAGCCAAAw
401 GCATAGATGT AGAGTTCGCA AACAACCGGT ATGCCACTGA CTACGcCCAT
10 451 ACCCagGGAA CAAAAGGCC TTACCGTCGC CCTCAATGTC CCGGTTGTCC
501 AAGCTGCACA AAACCTCATA CAAGCAGCCC AAAATGTGGG CAAAAGTAAA
551 AATAAACGCG TTAATGCCAT GGCTGCAGCC AATGCTGCAT GGCAGAGTTA
601 TCAAGCAACC CAACAATGC AACAATTGTC TCCAAGCAGC AGTGGCGGAC
651 AAGGTCAAAA CTACAATCAA AGCCCCAGTA TCAGTGTGTC CATTAC.TAC
15 701 GCGGAACAGA AAAGTCGTAA CGAGCAAAAA AGACATTACA CCGAAgCGGC
751 AgCAAGTCAA ATTATCGGCA AAGGGCAAA CACACTTGCG GCAACAGGAA
801 GTGGGGAGCA GTCCAATATC AATATTACAG GTTCCGATGT CATCGGCCAT
851 GCAGGTACTC C.CTCATTGC CGACAACCAT ATCAGACTCC AATCTGCCAA
901 ACAGGACGGC AGCGAGCAAA GCAAAAACAA AAGCAGTGGT TGGAAATCGAG
20 951 GCGTACGTnn CAAAATAGGC AACGGCATCA GGTTTGGAAT TACCGCCGGA
1001 GGAATATCG GTAAAGGTAA AGAGCAAGGG GGAAGTACTA CCCACCGCCA
1051 CACCCATGTC GGCAGCACAA CCGGCAAAAC TACCATCCGA AGCGGCGGGG
1101 GATACACCCC TCAAAAGGTGT GCAGCTCATC GGCAAAAGGCA TACAGGCAGA
1151 TACGCGCAAC CTGCATATAG AAAGTGTTCa AGATACTGAA ACCTATCAGA
25 1201 GCAAAACAGCA AAACGGCAAT GTCCAAGTTt ACTGTCGGTT ACGGATTGAG
1251 TGCAAGCGGC AGTTACCGCC AAAGCAAAGT CAAAGCAGAC CATGCCTCCG
1301 TAACCGGGCA AAgCGGTATT TATGCCGGAG AAGACGGCTA TCAAAATyAAA
1351 GTyAGAGACA ACACAGACCT yAAGGGCGGT ATCATCACGT CTAGSCAAAG
1401 CGCAGAAGAT AAGGGCAAAA ACCTTTTTCa GACGGCCACC CTTACTGCCA
30 1451 GCGACATTCa AAACCACAGC CGCTACGAAG GCAGAAGCTT CCGCATAGGC
1501 GGCAGTTTCG ACCTGAACGG CGGCTGGGAC GGCACGGTTA CCGACAAACA
1551 AGGCAGGCCT ACCGACAGGA TAAGCCCGGC AGCCGGCTAC GGCAGCGACG
1601 GAGACAGCAA AAACAGCACC ACCCGCAGCG GCGTCAACAC CCACAACATA
1651 CACATCACCG ACGAAGCGGG ACAACTTGCC CGAACAGGCA GGACTGCAAA
35 1701 AGAAACCGAA GCGCGTATCT ACACCGGCAT CGACACCGAA ACTGCGGATC
1751 AACACTCAGG CCATCTGAAA AACAGCTTCG AC...

```

This corresponds to the amino acid sequence <SEQ ID 64; ORF116>:

```

1  ..RFIHDEAVGS NIGGGKMIVA AGQDINVRGX SLISDKGIVL KAGHDIDIST
51  AHNRYTGNEY HESXXSGVMG TGGLGFTIGN RKTDDTDRT NIVHTGSIIG
40 101 SLNGDTVTVA GNRYRQTGST VSSPEGRNTV TAKXIDVEFA NNRYATDYAH
151 TQEQKGLTVA LNVFVVQAAQ NFIQAAQNVG KSKNKRNVAM AAANAAWQSY
201 QATQOMQQFA PSSSAGQGQN YNQSPSISVS IXYGQKSRN EQKRHYTEAA
251 ASQIIGKQGT TLAATGSSEQ SNINITGSDV IGHAGTXLIA DNHIRLQSAK
301 QDGSEQSKNK SSGWNAQVRX KIGNGIRFGI TAGGNIGKKG EQGGSSTHRH
45 351 THVGSTTGKT TIRSGDITL KGVQLIGKI QADTRNLHIE SVQDTETYQS
401 KQONGNVQVT VGYGFSASGS YRQSKVKADH ASVTGQSGIY AGEDGYQIKV
451 RDNTDLKGGI ITSSQSAEDK GKNLFQTATL TASDIQNSR YEGRSFGIGG
501 SFDLNGGWDG TVTDKQGRPT DRISPAAGYG SDGDSKNSTT RSGVNTNINH
551 ITDEAGQLAR TGRKATEA RIYTGIDTET ADQHSGLKN SFD...

```

50 Computer analysis of this amino acid sequence gave the following results:

Homology with *pspA* putative secreted protein of *N.meningitidis* (accession number AF030941)

ORF116 and *pspA* protein show 38% aa identity in 502aa overlap:

```

Orf116: 6  EAVGSNIGGGKMIVAAGQDINVRGXSLISDKGIVLKAGHDIDISTAHNRYTGNEYHESXX 65
      +AV  + G ++I+ +G+DI V G ++I+D  +L A ++I + A R  E ++
55  PspA: 1 235 QAVSGTLDGKEIILVSGRDITVTGSNIADNHTILSAKNIVLKAETRSRSAEMNKKEK 1294
Orf116: 66  XXXXXXXXXXXXXXXNRKXXXXXXXXRTNIVHTGSIIGSLNGDTVTVAGNRYRQTGSTVSSPE 125
      ++K  + HT S++GSLNG+T+  AG Y QTGST+SSP+
60  PspA: 1295 SGLMGSGGIGFTAGSKKDTQNRSETVSHTESVVGSLNGNTLISAGKHYTQTGSTISSPQ 1354

```

Orf116: 126 GRNTVTAKXIDVEFANNRYATDYAHTQEQKGLTVALNVPXXXX---XXXXXXXXXXGKS 182
 G +++ I ++ A NRY+ + EQKG+TVA++VP GKS
 PspA: 1355 GDVGISSGKISIDAAQNRYSQESQVYEQKGVTVASVPVNTVMGAVDAVKAVQTVGKS 1414

5 Orf116: 183 KNKRVSXXXXXXXXXWQSYQATQOMQQA--PSSSAGQGQNYNQSPSISVSIXYGEQKSRN 240
 KN RV + + + A P +AGQG ISVS+ YGEQK+ +
 PspA: 1415 KNSRVNMAAANALNKGVD SGVALYNAARNPKKAAGQG-----ISVSVTYGEQKNTS 1466

10 Orf116: 241 EQKRHYTEAAASQIIGKQTTLAATSGEQSNINITGSDVIGHAGTXLIADNHRLQSAK 300
 E + T+ +I G G+ +L A+G+G+ S I ITGSDV G GT L A+N +++++A+
 PspA: 1467 ESRIKGTQVQEGKITGGGKVS LTASGAGKDSRITITGSDVYGKGT RLKAENAVQIEAAR 1526

15 Orf116: 301 QDGSEQSKNKSSGWNAGVRXKIGNGIRFGITAXXXXXXXXXXXXXSTTHRHVHGSTTGKT 360
 Q E+S+NKS+G+NAGV I GI FG TA T +R++H+GS +T
 PspA: 1527 QTHQERSENKSAGFNAGVAIAINKGISFGFTAGANYGKGYGNGDETAYRNSHIGSKDSQT 1586

20 Orf116: 361 TIRSGGDTTLKGVQLIGKGIQADTRNLHIESVQDTETYSKQONGNVQTVGYGFSASGS 420
 I SGGDT +KG QL GKG+ +LHIES+QDT ++ KQ+N + QTVGYGFS GS
 PspA: 1587 AIESGGDTVIKGGQLKGGKGVGTAE SLHIESLQDTAVFKGKQENVSAQTVGYGFSVGGG 1646

25 Orf116: 421 YRQSKVKADHASVTGQSGIYAGEDGYQIKVRDNTDLKGGIITSSQSAEDKGKNLPQTATL 480
 Y +SK +D+ASV QSGI+AG DGY+I+V T L G + S DK KNL +T+ +
 PspA: 1647 YNRKSSSDYASVNEQSGIFAGGDGYRIRVNGKTGLVGAAVSD---ADKSKNLLKTSEI 1703

Orf116: 481 TASDIQNHRSRYEGRSFGIGGSF 502
 DIQNH+ + G+ G F
 PspA: 1704 WHKDIQNHASAAASALGLSGGF 1725

Based on homology with *pspA*, it is predicted that this protein from *N.meningitidis*, and its epitopes, could be useful antigens for vaccines or diagnostics.

30 Example 15

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 65>

35 1 ..ACGACCGGCA GCCTCGGCGG CATACTGGCC GGCGGCGGCA CTCCTCTGCG
 51 CGCACCGTAT TTGGACAAAG CGGCGGAAAA CCTCGGTCCG GCGGGCAAAG
 101 CGGCGGTCAA CGCACTGGGC GGTGCGGCCA TCGGCTATGC AACTGGTGGT
 151 AGTGGTGGTG CTGTGGTGGG TCGGAATGTA GATTGGAACA ATAGGCAGCT
 201 GCATCCGAAA GAAATGGCGT TGGCCGACAA ATATGCCGAA GCCCTCAAGC
 251 GCGAAGTTGA AAAACGCGAA GGCAGAAAA TCAGCAGCCA AGAAGCGGCA
 301 ATGAGAATCC GCAGGCAGAT ATCGTTGGG TGGACAAAGG TTCCCAAGAC
 351 GGCTATACCG ACCAAAGCGT CATATCCCTT ATCGGAATGA

40 This corresponds to the amino acid sequence <SEQ ID 66; ORF118>:

1 ..TTGSLGGILA GGGTSLAAPY LDKAAENLGP AGKAAVNALG GAAIGYATGG
 51 SGGAVVGANV DWNNRQLHPK EMALADKYAE ALKREVEKRE GRKISSQEAA
 101 MRIRRQICVG WTKVPKTAIP TKASYPLSE*

Computer analysis of this amino acid sequence reveals two putative transmembrane domains.

45 Based on this analysis, it is predicted that this protein from *N.meningitidis*, and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 16

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 67>

1 ..CAATGCCGTC TGAAAAGCTC ACAATTTTAC AGACGGCATT TGTTATGCAA

51 GTACATATAC AGATTCCCTA TATACTGCC AGrkGCGTGC GTgGCTGAAG
 101 ACACCCCTTA CGCTTGCTAT TTGrAACAGC TCCAAGTCAC CAAAGACGTC
 151 AACTGGAACC AGGTACwACT GCGGTACGAC AAATGGGACT ATAAACAGGA
 201 AGGCTTAACC GGAGCCGGAG CAGCGATTAT TGGCTGGCT GTTACCGTGG
 5 251 TFACTGCGGG CGCGGGAgCC GGAGCCGAC TGGGcTTAA CGGCGCGGCc
 301 GCAGCGGCAA CCGATGCCGC ATTCCGCTCG CTGGCCAGCC AGGcTTCCGT
 351 ATCGCTCATC AaCAACAAAG GCAATATCGG TAaCACCCCTG AAAGAGCTGG
 401 GCAGAAGCAG CACGGTGAAG AATCTGATGG TTGCCGTCGc tACCGCagGC
 451 GTagCggaCA AAATCGGTGC TTCGGCACTG AACAAATGTCA GCGATAAGCA
 10 501 GTGGATCAAC AACCTGACCG TCAACCTGGC CAATGCGGGC AGTGCCGCAC
 551 TGATTAATAC CGCTGTCAAC GCGCGCAGCc tgAAAGACAA TCTGGAAGCG
 601 AATATCCTTG CGGCTTTGGT GAATACTGCG CATGGAGAAG CAGCCAGTAA
 651 AATCAAACAG TTGGATCAGC ACTACATTAC CCACAAGATT GCCCaTGCCA
 701 TAGCGGGCTG TGGGcTGCG GCGGCGAATA AGGGCAAGTG TCAGGATGGT
 15 751 GCGATagGTG CGGCTGTGGG CGAGATAGTC GGGGagGCTT TGACAAACGG
 801 CAAAAATCCT GACACTTTGA CAGCTAAAgA ACGCGaACAG ATTTTGGCAT
 851 ACAGCAAAC TGGTTGCGGT ACGGTAAGCG GTGTGGTCGG CGGCGATGTA
 901 AATGCGGCGG CGAATGCGGC TGAGGTAGCG GTGAAAATA ATCAGCTTAG
 951 CGACAAATGA

20 This corresponds to the amino acid sequence <SEQ ID 68; ORF41>:

1 ..QCRLKSSQFY RRHLLCKYIY RFPIYCPXAC VAEDTPYACY LXQLQVTKDV
 51 NWNQVXLAYD KWDYKQEGLT GAGAAIALA VTVVVTAGAGA GAALGLNGAA
 101 AAATDAEFAS LASQASVSLI NKNKNIGNTL KELGRSSTVK NLMVAVATAG
 151 VADKIGASAL NNVSDKQWIN NLTVNLANAG SAALINTAVN GGLKDNLEA
 201 NILAALVNTA HGEAASKIKQ LDQHYITHKI AHAIAGCAAA AANKGKCQDG
 25 251 AIGAAVGEIV GEALTNGKNP DTLTAKEREQ ILAYSKLVAG TVSGVVGGDV
 301 NAAANAEEVA VKNNQLSDK*

Further work revealed the complete nucleotide sequence <SEQ ID 69>:

1 ATGCAAGTAA ATATTCAGAT TCCCTATATA CTGCCAGAT GCGTGGTGC
 30 51 TGAAGACACC CCCTACGCTT GCTATTTGAA ACAGCTCCAA GTCACCAAAG
 101 ACGTCAACTG GAACCAGGTA CAACTGGCGT ACGACAAATG GGACTATAAA
 151 CAGGAAGGCT TAACCGGAGC CGGAGCAGCG ATTATTGCGC TGGCTGTTAC
 201 CGTGGTTACT GCGGCGCGG GAGCGGAGC CGCACTGGGC TTAAACGGCG
 251 CGGCCGACG GGCACCGAT GCCGATTTCG CCTCGCTGGC CAGCCAGGCT
 35 301 TCCGTATCGC TCATCAACAA CAAAGGCAAT ATCGGTAAAC CCCTGAAAGA
 351 GCTGGGCAGA AGCAGCACCG TGAAAAATCT GATGGTTGCC GTCGCTACCG
 401 CAGGCGTAGC CGACAAATC GGTGCTTCG CACTGAACAA TGTCAGCGAT
 451 AAGCAGTGA TCAACAACCT GACCGTCAAC CTGGCCAATG CGGGCAGTGC
 501 CGCACTGATT AATACCGCTG TCAACGGCGG CAGCCTGAAA GACAATCTGG
 40 551 AAGCGAATAT CCTTGGCGGT TTGGTGAATA CTGCGCATGG AGAAGCAGCC
 601 AGTAAATCA AACAGTTGGA TCAGCACTAC ATTACCCACA AGATTGCCCCA
 651 TGCCATAGCG GGCTGTGCGG CTGCGGCGGC GAATAAGGGC AAGTGTCAAG
 701 ATGGTGCGAT AGGTGCGGCT GTGGGCGAGA TAGTCGGGGA GGCTTTGACA
 751 AACGGCAAAA ATCTTGACAC TTTGACAGCT AAAGAACGCG AACAGATTTT
 45 801 GGCATACAGC AAATGCTTG CCGGTACGGT AAGCGGTGTG GTCGGCGGCG
 851 ATGTAATATG GCGGCGAAT GCGGCTGAGG TAGCGGTGAA AAATAATCAG
 901 CTTAGCGACA AAGAGGCTAG AGAATTGAT AACGAAATGA CTGCATGCGC
 951 CAAACAGAAT AATCTCAAC TGTGCAGAAA AAATACTGTA AAAAAGTATC
 1001 AAAATGTTGC TGATAAAAGA CTTGCTGCTT CGATTGCAAT ATGTACGGAT
 50 1051 ATATCCCCTA GACTGAATG TAGAACAATC AGAAAACAAC ATTTGATCGA
 1101 TAGTAGAAGC CTTATTTCAT CTTGGGAAGC AGGTCTAATT GGTAAAGATG
 1151 ATGAATGGTA TAAATTATTC AGCAATCTT ACACCCAAGC AGATTTGGCT
 1201 TTACAGTCTT ATCATTTGAA TACTGCTGCT AAATCTTGGC TTCAATCGGG
 1251 CAATACAAAG CCTTTATCCG AATGGATGTC CGACCAAGGT TATACACTTA
 55 1301 TTTCAAGAGT TAATCCTAGA TTCATTCCAA TACCAAGAGG GTTTGTAAAA
 1351 CAAAATACAC CTATTACTAA TGTCAAATAC CCGGAAGGCA TCAGTTTCGA
 1401 TACAAACCTA AAAAGACATC TGGCAAATGC TGATGGTTTT AGTCAAAAAC
 1451 AGGGCATTAA AGGAGCCCAT AACCGCACCA ATTTTATGGC AGAACTAAAT
 1501 TCACGAGGAG GACCGGTAAT ATCTGAAACC CAACTGATA TTGAAGGCAT
 60 1551 TACCCGAATT AAATATGAGA TTCCTACACT AGACAGGACA GGTAAACCTG
 1601 ATGGTGGATT TAAGGAAATT TCAAGTATAA AAATGTTTA TAATCCTAAA
 1651 AAATTTCTG ATGATAAAAT ACTTCAATG GCTCAAAATG CTGCTTCACA
 1701 AGGATATTCA AAAGCCTCTA AAATTGCTCA AAATGAAAGA ACTAAATCAA
 1751 TATCGGAAAG AAAAATGTG ATTCAATTCT CAGAAACCTT TGACGGAATC
 65 1801 AAATTTAGAT CATATTTTGA TGTAAATACA GGAAGAATTA CAAACATTCA
 1851 CCCAGAATAA

15 Computer analysis of this amino acid sequence predicts a transmembrane domain, and homology with an ORF from *N.meningitidis* (strain A) was also found.

		10	20	30	40	50	60	69
20	orf41.pep	YRRHLLCKYIYRFPYICPXACVAEDTPYACYLXQLQVTKDWNWNVQVXLAYDKWDYKQEG						
	orf41a				: ::			
					YLKQLQVAKNINWNVQVQLAYDRWDYKQEG			
					10	20	30	
25		70	80	90	100	110	120	129
	orf41.pep	TGAGAAIIALAVTVVTAGAGAGAALGLNGAAAAATDAAFASLASQASVSLINNKG						
	orf41a	TEAGAAIIALAVTVVTSGAGTGAVLGLNGAXAAATDAAFASLASQASVSFINNKG						
		40	50	60	70	80	90	
30		130	140	150	160	170	180	189
	orf41.pep	LKELGRSSTVKNLMVAVATAGVADKIGASALNNVSDKQWNNLTVNLANAGS						
	orf41a	LKELGRSSTVKNLVVAAATAGVADKIGASALXNVSDKQWNNLTVNLANAGS						
35			100	110	120	130	140	150
		190	200	210	220	230	240	249
	orf41.pep	NGGSLKDNLEANILAALVNTAHGEAASKIKQLDQHYITHKIAHAIAGCAAAA						
	orf41a	NGGSLKDXLEANILAALVNTAHGEAASKIKQLDQHYIVHKIAHAIAGCAAAA						
40			160	170	180	190	200	210
		250	260	270	280	290	300	309
	orf41.pep	GAIGAAGVEIGEALTNKGNPDTLTAKEREQILAYSKLVAGTVSGVVGDDV						
45	orf41a	GAIGAAGVEIGEALTNKGNPDTLTAKEREQILAYSKLVAGTVSGVVGDDV						
		220	230	240	250	260	270	
50		310	320					
	orf41.pep	AVKNNQLSDKX						
	orf41a	AVKNNQLSDXEGREFDNEMTACAKQNXPOLCRKNTVKKYQNVDKRLAASIAICTDIS						
		280	290	300	310	320	330	

55 1 ..TATCTGAAAC AGCTCCAAGT AGCGAAAAAC ATCAACTGGA ATCAGGTGCA
 51 GCTTGCTTAC GACAGATGGG ACTACAAACA GGAGGGCTTA ACCGAAGCAG
 101 GTGCGGCGAT TATCGCACTG GCCGTTACCG TGGTCACTC AGGCGCAGGA
 151 ACCGGAGCCG TATTGGGATT AAACGGTGCG NCCGCCGCCG CAACCGATGC

201 AGCATTGCGC TCTTTGGCCA GCCAGGCTTC CGTATCGTTC ATCAACAACA
 251 AAGGCGATGT CGGCAAAACC CTGAAAGAGC TGGGCAGAAG CAGCACGGTG
 301 AAAAATCTGG TGGTTGCCGC CGCTACCGCA GCGGTAGCCG ACAAATCGG
 351 CGCTTCGGCA CTGANCAATG TCAGCGATAA GCAGTGGATC AACAACTGA
 5 401 CCGTCAACCT AGCCAATGCG GGCAGTCCG CACTGATTAA TACCGCTGTC
 451 AACGGCGGCA GCGTGAAAGA CANTCTGGAA GCGAATATCC TTGCGGCTTT
 501 GGTCAATACC GCGCATGGAG AAGCAGCCAG TAAATCAAA CAGTTGGATC
 551 AGCACTACAT AGTCCACAAG ATTGCCCATG CCATAGCGGG CTGTGCGGCA
 601 GCGGCGGCGA ATAAGGGCAA GTGT CAGGAT GGTGCGATAG GTGCGGCTGT
 10 651 GGGCGAGATA GTCGGGGAGG CTTTGACAAA CGGCAAAAT CCTGACACTT
 701 TGACAGCTAA AGAACGCGAA CAGATTTTGG CATACAGCAA ACTGGTTGCC
 751 GGTACGGTAA GCGGTGTGGT CGGCGGCGAT GTAAATGCGG CGGCGAATGC
 801 GCGTGAGGTA GCGGTGAAAA ATAATCAGCT TAGCGACNAA GAGGGTAGAG
 851 AATTTGATAA CGAAATGACT GCATGCGCCA AACAGAATAN TCCTCAACTG
 15 901 TGCAGAAAAA ATACTGTAAA AAAGTATCAA AATGTTGCTG ATAAAAGACT
 951 TGCTGCTTCG ATTGCAATAT GTACGGATAT ATCCCGTAGT ACTGAATGTA
 1001 GAACAATCAG AAAACAACAT TTGATCGATA GTAGAAGCTT TCATTATCT
 1051 TGGGAAGCAG GTCTAATTGG TAAAGATGAT GAATGGTATA AATTATTGAG
 1101 CAAATCTTAC ACCCAAGCAG ATTTGGCTTT ACAGTCTTAT CATTTGAATA
 20 1151 CTGCTGCTAA ATCTTGGCTT CAATCGGGCA ATACAAAGCC TTTATCCGAA
 1201 TGGATGTCCG ACCAAGGTTA TACACTTATT TCAGGAGTTA ATCTAGATT
 1251 CATTCCAATA CCAAGAGGGT TTGTAAACA AAATACACCT ATTACTAATG
 1301 TCAAATACCC GGAAGGCATC AGTTTCGATA CAAACCTANA AAGACATCTG
 1351 GCAAATGCTG ATGGTTTGTAG TCAAGAACAG GGCATTAAAG GAGCCATAA
 25 1401 CCGACCAAT NTTATGGCAG AACTAAATTC ACGAGGAGGA NGNGTAAAT
 1451 CTGAAACCCA NACTGATATT GAAGGCATTA CCGGAATTAA ATATGAGATT
 1501 CCTACACTAG ACAGGACAGG TAAACCTGAT GGTGGATTAA AGGAAATTC
 1551 AAGTATAAAA ACTGTTTATA ATCCTAAAAA NTTTNNNGAT GATAAATAC
 1601 TTCAAATGGC TCAANATGCT GNTTCACAAG GATATTCAA AGCCTCTAAA
 30 1651 ATTGCTCAA ATGAAAGAAC TAAATCAATA TCGGAAAGAA AAAATGTCAT
 1701 TCAATTCTCA GAAACCTTTG ACGGAATCAA ATTTAGANNN TATNTNGATG
 1751 TAAATACAGG AAGAATTACA AACATTACAC CAGAATAA

This encodes a protein having the partial amino acid sequence <SEQ ID 72>:

1 YLKQLQVAKN INWNQVQLAY DRWDYKQEGL TEAGAAIAL AVTVVTSAG
 35 51 TGAVLGLNGA XAAATDAFA SLASQASVSF INNKGDVGKT LKELGRSSTV
 101 KNLVVAAATA GVADKIGASA LXNVSDKQWI NNLTVNLANA GSAALINTAV
 151 NGGSLKDXLE ANILAALVNT AHGEAASKIK QLDQHYIVHK IAHAIAGCAA
 201 AAANKGKCQD GAIGAAVGEI VGEALTNGKN PDTLTAKERE QILAYSKLVA
 251 GTVSGVVGGD VNAANAAEV AVKNNQLSDX EGREFDNEMT ACAKONXPOL
 40 301 CRKNTVKKYQ NVADKRLAAS IAICTDISRS TECRTIRKQH LIDSRSLHSS
 351 WEAGLIGKDD EWYKLFKSKY TOADLALQSY HLNATAKSWL QSGNTKPLSE
 401 WMSDQGYTLI SGVNPFIPI PRGFVKQNTPT ITNVKYPEGI SFDTNLXRHL
 451 ANADGFSQEQ GIKGAHNRTN XMAELNSRGG XVKSETXTDI EGITRIKYEI
 501 PTLDRGTGKPD GGFKEISSIK TVYNPKFXFD DKILQMAQXA XSQGYSKASK
 45 551 IAQNERTKSI SERKNVIQFS ETFDGIKFRX YXDVNTGRIT NIHPE*

ORF41a and ORF41-1 show 94.8% identity in 595 aa overlap:

		10	20	30
orf41a.pep		YLKQLQVAKNINWNQVQLAYDRWDYKQEGLTEAGAA		
50 orf41-1		MQVNIQIPYILPRCVRAEDTPYACYLKQLQVTKDVNWNQVQLAYDKWDYKQEGLTGAGAA		
		10	20	30
		40	50	60
orf41a.pep		IIALAVTVVTSAGTGAVLGLNGAXAAATDAAFASLASQASVSFINNKGDVGKTLKELGR		
55 orf41-1		IIALAVTVVTAGAGAGAALGLNGAAAAATDAAFASLASQASVSLINNKNIGNTLKELGR		
		70	80	90
		100	110	120
60 orf41a.pep		SSTVKNLVVAAATAGVADKIGASALXNVSDKQWINNLTVNLANAGSAAALINTAVNGGSLK		
orf41-1		SSTVKNLVAVATAGVADKIGASALNNVSDKQWINNLTVNLANAGSAAALINTAVNGGSLK		
		130	140	150
		160	170	180
65 orf41a.pep		DXLEANILAALVNTAHGEAASKIKQLDQHYIVHKIAHAIAGCAAAAANKGKCQDGAIGAA		

-100-

	orff41-1	DNLEANI L AALVNTAHGEAASKIKQLDQHYITHKIAHAIAGCAAAAANKGKCQDGAIGAA
		190 200 210 220 230 240
5	orff41a.pep	VGEIVGEALTNGKNPDTLTAKEREQILAYSKL VAGTVSGVVGDDVNAAAANAEEVAVKNNQ
	orff41-1	VGEIVGEALTNGKNPDTLTAKEREQILAYSKL VAGTVSGVVGDDVNAAAANAEEVAVKNNQ
10		250 260 270 280 290 300
	orff41a.pep	LSDXEGREFDNEMTACAKQNXPOLCRKNTVKKYQNVADKRLAASIAICTDISRSTECRTI
	orff41-1	LSDXEGREFDNEMTACAKQNNPOLCRKNTVKKYQNVADKRLAASIAICTDISRSTECRTI
15		310 320 330 340 350 360
	orff41a.pep	RKQHLIDSRSLHSSWEAGLIGKDDWEYKLF SKSYTQADLALQSYHLNTAAKSWLQSGNTK
20	orff41-1	RKQHLIDSRSLHSSWEAGLIGKDDWEYKLF SKSYTQADLALQSYHLNTAAKSWLQSGNTK
		370 380 390 400 410 420
	orff41a.pep	PLSEWMSDQGYTLISGVNPRFIPIPRGFVKQNT PITNVKYPEGISFDTNLKRHLANADGF
25	orff41-1	PLSEWMSDQGYTLISGVNPRFIPIPRGFVKQNT PITNVKYPEGISFDTNLKRHLANADGF
		430 440 450 460 470 480
30	orff41a.pep	SQEQGIKGAHNRTNFXMAELNSRGGXVKSETXTDIEGITRIKYEIPTLDRTGKPDGGFKEI
	orff41-1	SQKQGIKGAHNRTNFXMAELNSRGGXVKSETXTDIEGITRIKYEIPTLDRTGKPDGGFKEI
		490 500 510 520 530 540
35	orff41a.pep	SSIKTVYNPKKFXDDKILQMAQXAXSQGYSKASKIAQNERTKSI SERKNVQFSETFDGI
	orff41-1	SSIKTVYNPKKFXDDKILQMAQXAXSQGYSKASKIAQNERTKSI SERKNVQFSETFDGI
40		550 560 570 580 590 600
	orff41a.pep	KFRXYXDVNTGRITNIHPEX
45	orff41-1	KFRSYFDVNTGRITNIHPEX
		610 620

Amino acids 25-619 of ORF41-1 were amplified as described above. Figure 6 shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF41-1.

Based on this analysis, it is predicted that this protein from *N.meningitidis*, and its epitopes, could be useful antigens for vaccines or diagnostics.

50 Example 17

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 73>

55

```

1  ATGGCAATCA TTACATTGTA TTATTCTGTC AATGGTATTT TAAATGTATG
51  TGCAAAAGCA AAAAATATTC AAGTAGTTGC CAATAATAAG AATATGGTTC
101 TTTTGGGTT TTTGGsmrGC ATCATCGGCG GTTCAACCAA TGCCATGTCT
151 CCCATATTGT TAATATTTTT GCTTAGCGAA ACAGAAAATA AAAATcgTAT
201 CGTAAAATCA AGCAATCTAT GCTATCTTTT GGCGAAAATT GTTCAAATAT
251 ATATGCTAAG AGACCAATAT TGGTTATTAA ATAAGAGTGA ATACGdTTTA
301 ATATTTTAC TGCCGTATT GTCTGTATT GGATTGTATG TTGGAATTCG
351 GTTAAGGACT AAGATTAGCC CaaATTTTTT TAAAATGTTA ATTTTTATTG

```

401 tTTTATTGGT ATTGGCTCTG AAAATCGGGC AttCGGGTTT AatCAAACCTT
451 TAA

This corresponds to the amino acid sequence <SEQ ID 74; ORF51>:

5 1 MAIITLYYSV NGILNCAKA KNIQVVANNK NMVLFGLXX IGGSTNAMS
51 PILLIFLLSE TENKNRIVKS SNLCYLLAKI VQIYMLRDQY WLLNKSEYXL
101 IFLLSVLSVI GLYVGIRLRT KISPNEFKML IFIVLLVLAL KIGHSGLIKL
151 *

Further work revealed the complete nucleotide sequence <SEQ ID 75>:

10 1 ATGCAAGAAA TAATGCAATC TATCGTTTTT GTTGCTGCCG CAATACTGCA
51 CGGAATTACA GGCATGGGAT TTCCGATGCT CGGTACAACC GCATTGGCTT
101 TTATCATGCC ATTGTCTAAG GTTGTGCCT TGGTGGCATT ACCAAGCCTG
151 TTAATGAGCT TGTGGTTCT ATGCAGCAAT AACAAAAAGG GTTTTTGGCA
201 AGAGATTGTT TATTATTAA AACCTATAA ATTGCTTGCT ATCGGCAGCG
15 251 TCGTTGGCAG CATTITGGGG GTGAAGTGC TTTTGATACT TCCAGTGTCT
301 TGGCTGCTTT TACTGATGGC AATCATTACA TTGTATTATT CTGTCAATGG
351 TATTTTAAAT GTATGTGCAA AAGCAAAAAA TATTCAAGTA GTTGCCATA
401 ATAAGAATAT GGTCTTTTTT GGGTTTTTGG CAGGCATCAT CGGCGGTTCA
451 ACCAATGCCA TGTCTCCCAT ATTGTTAATA TTTTGTCTTA GCGAAACAGA
501 AAATAAAAT CGTATCGTAA AATCAAGCAA TCTATGCTAT CTTTGGCGA
20 551 AAATTGTCA AATATATATG CTAAGAGACC AGTATTGGT ATTAATAAG
601 AGTGAATACG GTTAATATT TTTACTGTCC GTATTGTCTG TTATTGGATT
651 GTATGTTGGA ATTCGGTTAA GACTAAGAT TAGCCCAAT TTTTAAAAA
701 TGTTAATTTT TATTGTTTAA TTGGTATTGG CTCTGAAAT CGGGCATTCTG
751 GGTTAATCA AACTTTAA

25 This corresponds to the amino acid sequence <SEQ ID 76; ORF51-1>:

1 MQEIMQSIVF VAAAILHGIT GMGFPMGLTT ALAFIMPLSK VVALVALPSL
51 LMSLLVLCN NKKGFQWQIV YYLKTYKLLA IGSVVGSIIG VKLLILPVS
101 WLLLLMAIIT LYYSVNGILN VCAKAKNIQV VANNKNMVLF GFLAGIIGGS
151 TNAMSPILLI FLLSETENKN RIVKSSNLCY LLAKIVQIYM LRDQYWLKN
30 201 SEYGLIFLLS VLSVIGLYVG IRLRTKISPN FFKMLIFIVL LVLALKIGHS
251 GLIKL*

Computer analysis of this amino acid sequence reveals three putative transmembrane domains. A

corresponding ORF from strain A of *N.meningitidis* was also identified:

Homology with a predicted ORF from *N.meningitidis* (strain A)

35 ORF51 shows 96.7% identity over a 150aa overlap with an ORF (ORF51a) from strain A of *N.meningitidis*:

				10	20	30
orf51.pep				MAIITLYYSV	NGILNCAKA	KNIQVVANNK
40 orf51a	YKLLAIGSVVGSILGVKLLILPVS	WLLLLMAIITLYYSV	NGILNCAKA	KNIQVVANNK		
	80	90	100	110	120	130
		40	50	60	70	80
orf51.pep	NMVLFGLXXI	IGGSTNAMSPILLIFLL	SETENKNRIVKSSNLCYLLAKIVQIYMLRDQY			
45 orf51a	NMVLFGLAGI	IGGSTNAMSPILLIFLL	SETENKNRIAKSSNLCYLLAKIVQIYMLRDQY			
	140	150	160	170	180	190
	100	110	120	130	140	150
50 orf51.pep	WLLNKSEYXLI	FLLSVLSVIGLYVGIRLRTKISPNEFKMLIFIVLLVLALKIGHSGLIKL				
orf51a	WLLNKSEYGLI	FLLSVLSVIGLYVGIRLRTKISPNEFKMLIFIVLLVLALKIGYSGLIKL				
	200	210	220	230	240	250

ORF51-1 and ORF51a show 99.2% identity in 255 aa overlap:

```

5  orf51a.pep  MQEIMQSIVFVAAAILHGITGMGFPM LGTTALAFIMPLSKVVALVALPSLLMSLLVLCSN
   orf51-1    MQEIMQSIVFVAAAILHGITGMGFPM LGTTALAFIMPLSKVVALVALPSLLMSLLVLCSN

10 orf51a.pep  NKKGFQWEIVYYLKTYKLLAIGSVVGSILGVKLLILPVSLLLLLMAIITLYYSVNGILN
   orf51-1    NKKGFQWEIVYYLKTYKLLAIGSVVGSILGVKLLILPVSLLLLLMAIITLYYSVNGILN

15 orf51a.pep  VCAKAKNIQVVANNKNMVLFGFLAGIIGGSTNAMS PILLIFLLSETENKNRIAKSSNL CY
   orf51-1    VCAKAKNIQVVANNKNMVLFGFLAGIIGGSTNAMS PILLIFLLSETENKNRIVKSSNL CY

20 orf51a.pep  LLAKIVQIYMLRDQYWLLNKSEYGLIFLLSVLSVIGLYVGIRLR TKISPNFFKMLIFIVL
   orf51-1    LLAKIVQIYMLRDQYWLLNKSEYGLIFLLSVLSVIGLYVGIRLR TKISPNFFKMLIFIVL

   orf51a.pep  LVLALKIGYSGLIKLX
   orf51-1    LVLALKIGHSGLIK LX

```

The complete length ORF51a nucleotide sequence <SEQ ID 77> is:

```

1  ATGCAAGAAA TAATGCAATC TATCGTTTTT GTTGCTGCCG CAATACTGCA
51 CGGAATTACA GGCATGGGAT TTCCGATGCT CGGTACAACC GCATTGGCTT
25 101 TTATCATGCC ATTGTCTAAG GTTGTGCGCT TGGTGGCATT ACCAAGCCTG
   151 TTAATGAGCT TGTGGTTCTT ATGCAGCAAT AACAAAAAGG GTTTTGGGCA
   201 AGAGATTGTT TATTATTTAA AAACCTATAA ATTGCTTGCT ATCGGCAGCG
   251 TCGTTGGCAG CATTTTGGGG GTGAAGTTGC TTTTGATACT TCCAGTGTCT
   301 TGGCTGCTTT TACTGATGGC AATCATTACA TTGTATTATT CTGTCAATGG
   351 TATTTTAAAT GTATGTGCAA AAGCAAAAAA TATTCAAGTA GTTGCCAATA
30 401 ATAAGAATAT GGTTCCTTTT GGGTTTITGG CAGGCATCAT CGGCGGTTCA
   451 ACCAATGCCA TGTCTCCCAT ATTGTTAATA TTTTGTCTTA GCGAAACAGA
   501 GAATAAAAAT CGTATCGCAA AATCAAGCAA TCTATGCTAT CTTTGGCAA
   551 AAATTGTTCA AATATATATG CTAAGAGACC AGTATGGTT ATTAAATAAG
60 601 AGTGAATACG GTTTAATATT TTTACTGTCC GTATTGTCTG TTATTGGATT
   651 GTATGTTGGA ATTCGGTTAA GGAATAAGAT TAGCCCAAAT TTTTAAATAA
35 701 TGTTAATTTT TATTGTTTAA TTGGTATTCG CTCTGAAAT CGGGTATTCA
   751 GGTTTAATCA AACTTTAA

```

This encodes a protein having amino acid sequence <SEQ ID 78>:

```

40 1  MQEIMQSIVF VAAAILHGIT GMGFPM LGTT ALAFIMPLSK VVALVALPSL
   51  IMSLLVLCSN NKKGFQWEIV YYLKTYKLLA IGSVVGSILG VKLLILPVS
   101 WLLLMIAIT LYYSVNGILN VCAKAKNIQV VANNKNMVLFG FLAGIIGGS
   151 TNAMSPILLI FLLSETENKN RIAKSSNL CY LLAKIVQIYM LRDQYWLLNK
   201 SEYGLIFLLS VLSVIGLYVG IRLR TKISPN FFKMLIFIVL LVLALKIGYS
   251 GLIKL*

```

45 Based on this analysis, it is predicted that this protein from *N.meningitidis*, and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 18

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 79>

```

50 1  ATGAGACATA TGAAATACAA AAATTATTTA CTAGTATTTA TAGTTTTACA
   51  TATAGCCTTG ATAGTAATTA ATATAGTGTT TGGTTATTTT GTTTTCTAT
   101 TTGATTTTTT TCGGTTTTTG TTTTTGCAA ACGTCTTTCT TGCTGTAAT
   151 TTATTATTTT TAGAAAAAAA CATAAAAAAC AAATTATTGT TTTTATTGCC
   201 GATTTCTATT ATTATATGGA TGGTAATTCA TATTAGTATG ATAAATATAA
55 251 AATTTTATAA ATTTGAGCAT CAAATAAAGG AACAAAATAT ATCCTCGATT
   301 ACTGGGGTGA TAAACCACA TGATAGTTAT AATTATGTTT ATGACTCAAA

```


351 TGGATATGCT AAATTAAAAG ATAATCATAG ATATGGTAGG GTAATTAGAG
 401 AAACACCTTA TATTGATGTA GTTGCATCTG ATGTTAAAAA TAAATCCATA
 451 AGATTAAGCT TGGTTTGTGG TATTCATTCA TATGCTCCAT GTGCCAATTT
 501 TATAAAATTT GTCAGG..

5 This corresponds to the amino acid sequence <SEQ ID 80; ORF82>:

1 MRHMKIONYL LVFIVLHIAL IVINIVFGYF VFLDFFAFL FFANVFLAVN
 51 LLFLEKNIKN KLLFLLPISI IIWMVIHISM INIKFYKFEH QIKEQNISSI
 101 TGVIKPHDSY NYVYDSNGYA KLKDNHRYGR VIRETPYIDV VASDVKNKSI
 151 RLSLVCGIHS YAPCANFIKF VR..

10 Further work revealed the complete nucleotide sequence <SEQ ID 81>:

1 ATGAGACATA TGAAAATAA AAATTATTTA CTAGTATTTA TAGTTTTACA
 51 TATAGCCTTG ATAGTAATTA ATATAGTGTG TGGTTATTTT GTTTTTCTAT
 101 TTGATTTTTT TCGGTTTTTG TTTTGTGCAA ACGTCTTTCT TGCTGTAAT
 151 TTATTATTTT TAGAAAAAAA CATAAAAAAC AAATTATGTG TTTTATTGCC
 15 GATTTCTATT ATTATATGGA TGGTAATTCA TATTAGTATG ATAAATATAA
 251 AATTTTATAA ATTTGAGCAT CAAATAAAGG AACAAAATAT ATCCTCGATT
 301 ACTGGGGTGA TAAACCACA TGATAGTTAT AATTATGTTT ATGACTCAA
 351 TGGATATGCT AAATTAAAAG ATAATCATAG ATATGGTAGG GTAATTAGAG
 401 AAACACCTTA TATTGATGTA GTTGCATCTG ATGTTAAAAA TAAATCCATA
 20 451 AGATTAAGCT TGGTTTGTGG TATTCATTCA TATGCTCCAT GTGCCAATTT
 501 TATAAAATTT GCAAAAAAAC CTGTTAAAT TTATTTTAT AATCAACCTC
 551 AAGGAGATTT TATAGATAAT GTAATATTTG AAATTAATGA TGGAAACAAA
 601 AGTTTGIAC TGTAGATAA GTATAAACCA TTTTCTTCTA TTGAAAACAG
 651 TGTGTGATC GTATTAATTA TTTTATATTT AAAATTTAAT TTGCTTTTAT
 25 701 ATAGGACTTA CTTCAATGAG TTGGAATAG

This corresponds to the amino acid sequence <SEQ ID 82; ORF82-1>:

1 MRHMKKNKYL LVFIVLHIAL IVINIVFGYF VFLDFFAFL FFANVFLAVN
 51 LLFLEKNIKN KLLFLLPISI IIWMVIHISM INIKFYKFEH QIKEQNISSI
 101 TGVIKPHDSY NYVYDSNGYA KLKDNHRYGR VIRETPYIDV VASDVKNKSI
 30 151 RLSLVCGIHS YAPCANFIKF AKKPVKIIFY NQPQGFIDN VIFEINDGNK
 201 SLVLLDKYKT FFLIENSVCI VLIIYLKFN LLLYRTYFNE LE*

Computer analysis of this amino acid sequence reveals a predicted leader peptide.

A corresponding ORF from strain A of *N.meningitidis* was also identified:

Homology with a predicted ORF from *N.meningitidis* (strain A)

35 ORF82 shows 97.1% identity over a 172aa overlap with an ORF (ORF82a) from strain A of *N.meningitidis*:

		10	20	30	40	50	60
orf82.pep		MRHMKIONYLLVFIVLHIALIVINIVFGYFVFLDFFAFLFFANVFLAVNLLFLEKNIKN					
40 orf82a		MRHMKKNKYL LVFIVLHIALIVINIVFGYFVFLDFFAFLFFANVFLAVNLLFLEKNIKN					
		10	20	30	40	50	60
		70	80	90	100	110	120
orf82.pep		KLLFLLPISIIWMVIHISM INIKFYKFEHQIKEQNISSITGVIKPHDSYNYVYDSNGYA					
45 orf82a		KLLFLLPISIIWMVIHISM INIKFYKFEHQIKEQNISSITGVIKPHDSYNYVYDSNGYA					
		70	80	90	100	110	120
		130	140	150	160	170	
50 orf82.pep		KLKDNHRYGRVIRETPYIDVVASDVKNKSIRLSLVCGIHSYAPCANFIKFVR					
orf82a		KLKDNHRYGRVIRETPYIDVVASDVKNKSIRLSLVCGIHSYAPCANFIKFAKKPVKIIFY					
		130	140	150	160	170	180

ORF82a and ORF82-1 show 99.2% identity in 242 aa overlap:

```

5  orf82a.pep  MRHMKKNKYNLLVFIVLHITLIVINIVFGYFVFLDFFAFLFFANVFLAVNLLFLEKNIKN
   orf82-1    MRHMKKNKYNLLVFIVLHIALIVINIVFGYFVFLDFFAFLFFANVFLAVNLLFLEKNIKN

10 orf82a.pep  KLLFLLPISIIWMVIHISMINIKFYKFEHQIKEQNISSITGVIKPHDSYNYVYDSNGYA
   orf82-1    KLLFLLPISIIWMVIHISMINIKFYKFEHQIKEQNISSITGVIKPHDSYNYVYDSNGYA

15 orf82a.pep  KLKDNHRYGRVIRETPYIDVVASDVKNKSIRLSLVCGIHSYAPCANFIKFAKKPVKIYFY
   orf82-1    KLKDNHRYGRVIRETPYIDVVASDVKNKSIRLSLVCGIHSYAPCANFIKFAKKPVKIYFY

   orf82a.pep  NQPQGD FIDNVIFEINDGKKS LYLDDKYKTF FLIENSVCIVLIILYLKFNLLLYRTYFNE
   orf82-1    NQPQGD FIDNVIFEINDGKKS LYLDDKYKTF FLIENSVCIVLIILYLKFNLLLYRTYFNE

20 orf82a.pep  LEX
   orf82-1    LEX

```

The complete length ORF82a nucleotide sequence <SEQ ID 83> is:

```

25 1  ATGAGACATA  TGAAAAATAA  AAATTATTTA  CTAGTATTTA  TAGTTTACAC
   51  TATAACCTTG  ATAGTAATTA  ATATAGTGTT  TGGTTATTTT  GTTTTCTAT
101  TTGATTTTTT  TGCCTTTTGT  TTTTGTGCAA  ACGTCTTCT  TGCTGTAAT
151  TTATTATTTT  TAGAAAAAAA  CATAAAAAAC  AAATTATTGT  TTTTATTGCC
201  GATTTCTATT  ATTATATGGA  TGGTAATTCA  TATTAGTATG  ATAAATATAA
251  AATTTTATAA  ATTTGAGCAT  CAAATAAAGG  AACAAAATAT  ATCCTCGATT
301  ACTGGGGTGA  TAAAACCACA  TGATAGTTAT  AATTATGTTT  ATGACTCAAA
351  TGGATATGCT  AAATTAAGAG  ATAATCATAG  ATATGGTAGG  GTAATTAGAG
401  AAACACCTTA  TATTGATGTA  GTTGCATCTG  ATGTAAAAAA  TAAATCCATA
451  AGATTAAGCT  TGGTTTGTGG  TATTCATTCA  TATGCTCCAT  GTGCCAATTT
501  TATAAAATTT  GCAAAAAAAC  CTGTTAAAT  TTATTTTAT  AATCAACCTC
35  551  AAGGAGATTT  TATAGATAAT  GTAATATTTG  AAATTAATGA  TGGAAAAAAA
601  AGTTTGTACT  TGTAGATAAA  GTATAAAACA  TTTTCTCTTA  TTGAAAACAG
651  TGTGTGATC  GTATTAATTA  TTTTATATT  AAAATTTAAT  TTGCTTTT...
701  ATAGGACTTA  CTTCAATGAG  TTGGAATAG

```

This encodes a protein having amino acid sequence <SEQ ID 84>:

```

40 1  MRHMKKNKYNL  LVFIVLHITL  IVINIVFGYF  VFLDFFAFL  FFANVFLAVN
   51  LLFLEKNIKN  KLLFLLPISI  IIWMVIHISM  INIKFYKFEH  QIKEQNISSI
101  TGVIKPHDSY  NYVYDSNGYA  KLKDNHRYGR  VIRETPYIDV  VASDVKNKSI
151  RLSLVCGIHS  YAPCANFIKF  AKKPVKIYFY  NQPQGD FIDN  VIFEINDGKK
201  SLYLDDKYKT  FFLIENSVC  VLIILYLKFN  LLLRITYFNE  LE*

```

Based on this analysis, it is predicted that this protein from *N.meningitidis*, and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 19

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 85>

```

50 1  ..ACCCCAACA  GCGTGACCGT  CTTGCCGTCT  TTCGGCGGAT  TCGGGCGTAC
   51  CGGCGCGACC  ATCAATGCAG  CAGGCGGGGT  CGGCATGACT  GCCTTTTCGA
101  CAACCTTAAT  TTCCGTAGCC  GAGGGCGCGG  TTGTAGAGCT  GCAGGCGGTG
151  AGAGCCAAAG  CCGTCAATGC  AACCGCCGCT  TGCATTTTTA  CGGTCTTGAG
201  TAAGGACATT  TTCGATTTC  TTTTATTTT  CCGTTTTCAG  ACGGCTGACT
251  TCCGCCTGTA  TTTTCGCCAA  AGCCATGCCG  ACAGCGTGCG  CCTTGACTTC
301  ATATTTAAAA  GCTTCCGCGC  GTGCCAGTTC  CAGTTCGCGC  GCATAGTTTT
55  351  GAGCCGACAA  CAGCAGGGCT  TGCGCCTTGT  CGCGCTCCAT  CTTGTCGATG

```

```

401   ACCGCCTGCA GCTTCGCAAA TGCCGACTTG TAGCCTTGAT GGTGCGACAC
451   AGCCAAGCCC GTGCCGACAA GCGCGATAAT GGCAATCGGT TGCCAGTAAT
501   TCGCCAGCAG TTTCACGAGA TTCATTCTCG ACCTCCTGAC GCTTCACGCT
551   GA

```

5 This corresponds to the amino acid sequence <SEQ ID 86; ORF124>:

```

1   ..TPNSVTVLPS FGGFGRGTGAT INAAGVGMT AFSTTLISVA EGAVVELQAV
51  RAKAVNATAA CIFTVLSKDI FDFLFIFRFQ TADFRLYFRQ SHADSVRLDF
101 IFKSFRACQF QFARIVLSRQ QQGLRLVALH LVDDRLLQLRK CRLVALMVRH
151 SQARADKRDN GNRLPVIRQQ FHEIHSRPPD ASR*

```

10 Computer analysis of this amino acid sequence predicts a transmembrane domain.

Further work revealed the complete nucleotide sequence <SEQ ID 87>:

```

15 1   ATGACTGCCT TTTCGACAAC CTTAATTTCC GTAGCCGAGG GCGCGTTGT
    51  AGAGCTGCAG GCCGTGAGAG CCAAAGCCGT CAATGCAACC GCGCGTTGCA
    101 TTTTACGGT CTTGAGTAAG GACATTTTCG ATTTCTTTT TATTTTCCGT
    151 TTTTCAGCGG CTGACTTCG CCGTGTTCG CGCCAAAGC ATGCCGACAG
    201 CGTGCGCCTT GACTTCATAT TTTTAGCTT CCGCGCGTGC CAGTTCCAGT
    251 TCGCGCGCAT AGTTTTCGAGC CGACAACAGC AGGGCTTGC CTTGTGCGCG
    301 CTCCATCTTG TCGATGACCG CCGCTGCTT CGCAAATGCC GACTTGTAGC
    351 CTTGATGGTG CGACACAGCC AAGCCCGTGC CGACAAGCGC GATAATGGCA
20 401 ATCGGTTGCC AGTTATTCGC CAGCAGTTT ACGAGATTCA TTCTCGACCT
    451 CCTGACGCTT CACGCTGA

```

This corresponds to the amino acid sequence <SEQ ID 88; ORF124-1>:

```

25 1   MTAFTTLIS VAEGAVVELQ AVRAKAVNAT AACIFTVLSK DIFDFLFIFR
    51  FQTADFRLEF RQSHADSVRL DFIFFSFRAC QFQFARIVLS RQQQGLRLVA
    101 LHLVDDRLLL RKRLVALMV RHSQARADKR DNGNRLPVIR QPFHEIHSRP
    151 PDASR*

```

A corresponding ORF from strain A of *N.meningitidis* was also identified:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF124 shows 87.5% identity over a 152aa overlap with an ORF (ORF124a) from strain A of *N.meningitidis*:

```

30 orf124.pep   10    20    30    40    50    60
    TPNSVTVLPSFGGFGRTGATINAAGVGMTAFSTTLISVAEGAVVELQAVRAKAVNATAA
    orf124a           10    20    30
    MTAFTTLISVAEGALVELQAVMAKAVNTTAA

35 orf124.pep   70    80    90    100   110   120
    CIFTVLSKDIFDFLFIFRFQTADFRLYFRQSHADSVRLDFIFKSFRACQFQFARIVLSRQ
    orf124a           40    50    60    70    80    90
    CIFTVLSKDIFDFLFIFRFQTADFRLEFRQSHADGVRLDFIFFSFRTRLFQFAGVVLRSQ

40 orf124.pep   130   140   150   160   170   180
    QQGLRLVALHLVDDRLLQLRKRLVALMVRHSQARADKRDNGNRLPVIRQQFHEIHSRPPD
    orf124a           100   110   120   130   140   150
    QQGLRLVALHFLNDRLLLRKRLVALMVRHRTADKRDDGNRLPVIRQQFHEIHSRPPD

50 orf124.pep   ASRX
    orf124a      VX

```

ORF124a and ORF124-1 show 89.5% identity in 152 aa overlap:

orf124-1.pep MTAFSTTLISVAEGAVVELQAVRAKAVNATAACIFTVLSDKDIFDFLFIFRFQTADFRLEFF
 orf124a MTAFSTTLISVAEGALVELQAVMAKAVNTTAACIFTVLSDKDIFDFLFIFRFQTADFRLEFF
 5 orf124-1.pep RQSHADSVRLDFFFSFRACQFQFARIVLSRQQQGLRLVALHLVDDRLLLRKCRVALMV
 orf124a RQSHADGVRLDFFFSFRTRLFQFAGVLSRQQQGLRLVALHFLNDRLLLRKSRLVALMV
 10 orf124-1.pep RHSQARADKRDNGNRLPVIRQQFHEIHSRPPDASRX
 orf124a RHRQTRADKRDDGNRLPVIRQQFHEIHSRPPDVX

The complete length ORF124a nucleotide sequence <SEQ ID 89> is:

1 ATGACCGCCT TTTCGACAAC CTTAATTTC GTAGCCGAGG GCGCGCTTGT
 51 AGAGCTGCAA GCCGTGATGG CCAAAGCCGT CAATACAACC GCCGCCTGCA
 15 101 TTTTTACGGT CTTGAGTAAG GACATTTTCG ATTTCTCTTT TATTTTCCGT
 151 TTTCAGACGG CTGACTTCCG CCTGTTTTTT CGCCAAAGCC ATGCCGACGG
 201 CGTGCGCCTT GACTTCATAT TTTTTAGCTT CCGCACGCGC CTGTTCCAGT
 251 TCGCGGGCGT AGTTTGTAGC CGACAACAGC AGGGCTTGCG CCTTGTGCGG
 301 CTTCATTTTC TCAATGACCG CCTGCTGCTT CGCAAAGGCC GACTTGTAGC
 20 351 CTTGATGGTG CGACACCGCC AAACCCGTGC CGACAAGCGC GATGATGGCA
 401 ATCGGTTGCC AGTTATTGCG CAGCAGTTTC ACGAGATTCA TTCTCGACCT
 451 CCTGACGTTT GA

This encodes a protein having amino acid sequence <SEQ ID 90>:

1 MTAFSTTLIS VAEGALVELQ AVMAKAVNTT AACIFTVLSK DIFDFLFIFR
 25 51 FQTADFRLEFF RQSHADGVRL DFFFSFRTR LFQFAGVLS RQQQGLRLVA
 101 LHFLNDRLLL RKSRLVALMV RHRQTRADKR DDGNRLPVIR QQFHEIHSRP
 151 PDV*

ORF124-1 was amplified as described above. Figure 7 shows plots of hydrophilicity, antigenic
 30 index, and AMPHI regions for ORF124-1.

Based on this analysis, it is predicted that this protein from *N.meningitidis*, and its epitopes, could be useful antigens for vaccines or diagnostics.

It will be appreciated that the invention has been described by means of example only, and that
 35 modifications may be made whilst remaining within the spirit and scope of the invention.

TABLE I – PCR primers

ORF	Primer	Sequence	Restriction sites
ORF 38	Forward	CGCGGATCCCATATG-TCGCCGAAAATTCCGA	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TTTTGCCGCGTTAAAAGC	
ORF 40	Forward	CGCGGATCCCATATG-ACCGTGAAGACCGCC	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-CCACTGATAACCGACAGA	
ORF 41	Forward	CGCGGATCCCATATG-TATTGAAACAGCTCCAAG	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TTCTGGGTGAATGTTA	
ORF 44	Forward	GCGGATCCCATATG-GGCACGGACAACCCC	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-ACGTGGGGAACAGTCT	
ORF 51	Forward	GCGGATCCCATATG-AAAAATATTCAAGTAGTTGC	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-AAGTTTGATTAAACCCG	
ORF 52	Forward	CGCGGATCCCATATG-TGCCAACCGCAATCCG	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TTTTTCCAGTCCGGCA	
ORF 56	Forward	GCGGATCCCATATG-GTTATCGGAATATTACTCG	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-GGCTGCAGAAGCTGG	
ORF 69	Forward	CGCGGATCCCATATG-CGGACGTGGTTGGTTTT	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-ATATCTTCCGTTTTTTTCAC	
ORF 82	Forward	CGCGGATCCGCTAGC-GTAAATTATTATTTTAGAA	BamHI-NheI XhoI
	Reverse	CCCGCTCGAG-TTCCAACCTATTGAAGTA	
ORF 114	Forward	CGCGGATCCCATATG-AATAAAGGTTTACATCGCAT	BamHI-NheI XhoI
	Reverse	CCCGCTCGAG-AATCGCTGCACCGCT	
ORF 124	Forward	CGCGGATCCCATATG-ACTGCCTTTTCGACA	BamHI-NheI XhoI
	Reverse	CCCGCTCGAG-GCGTGAAGCGTCAGGA	

TABLE II – Cloning, expression and purification

ORF	PCR/cloning	His-fusion expression	GST-fusion expression	Purification
orf 38	+	+	+	His-fusion
orf 40	+	+	+	His-fusion
orf 41	+	n.d.	n.d.	
orf 44	+	+	+	His-fusion
orf 51	+	n.d.	n.d.	
orf 52	+	n.d.	+	GST-fusion
orf 56	+	n.d.	n.d.	
orf 69	+	n.d.	n.d.	
orf 82	+	n.d.	n.d.	
orf 114	+	n.d.	+	GST-fusion
orf 124	+	n.d.	n.d.	

CLAIMS

1. A protein comprising an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, and 6.
2. A nucleic acid molecule which encodes a protein according to claim 1.
- 5 3. A nucleic acid molecule according to claim 2, comprising a nucleotide sequence selected from the group consisting of SEQ IDs 1, 3, and 5.
4. A protein comprising an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, and 90.
- 10 5. A protein having 50% or greater sequence identity to a protein according to claim 4.
6. A protein comprising a fragment of an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, and 90.
7. An antibody which binds to a protein according to any one of claims 4 to 6.
- 15 8. A nucleic acid molecule which encodes a protein according to any one of claims 4 to 6.
9. A nucleic acid molecule according to claim 8, comprising a nucleotide sequence selected from the group consisting of SEQ IDs 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, and 89.
- 20 10. A nucleic acid molecule comprising a fragment of a nucleotide sequence selected from the group consisting of SEQ IDs 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, and 89.
11. A nucleic acid molecule comprising a nucleotide sequence complementary to a nucleic acid molecule according to any one of claims 8 to 10.

12. A nucleic acid molecule comprising a nucleotide sequences having 50% or greater sequence identity to a nucleic acid molecule according to any one of claims 8 to 11.
13. A nucleic acid molecule which can hybridise to a nucleic acid molecule according to any one of claims 8 to 12 under high stringency conditions.
- 5 14. A composition comprising a protein, a nucleic acid molecule, or an antibody according to any preceding claim.
15. A composition according to claim 14 being a vaccine composition or a diagnostic composition.
16. A composition according to claim 14 or claim 15 for use as a pharmaceutical.
- 10 17. The use of a composition according to claim 14 in the manufacture of a medicament for the treatment or prevention of infection due to Neisserial bacteria, particularly *Neisseria meningitidis*.

FIG. 1A

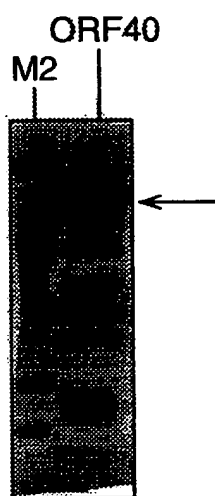


FIG. 1B

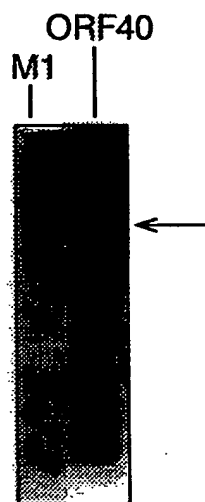
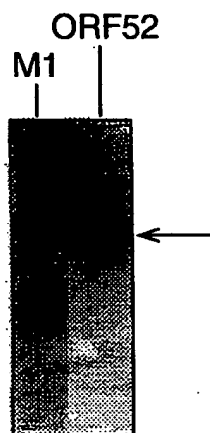
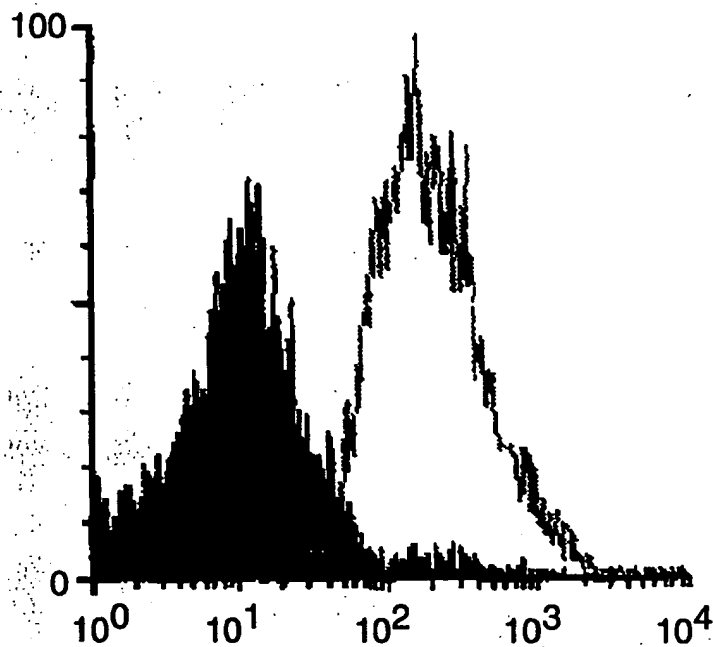
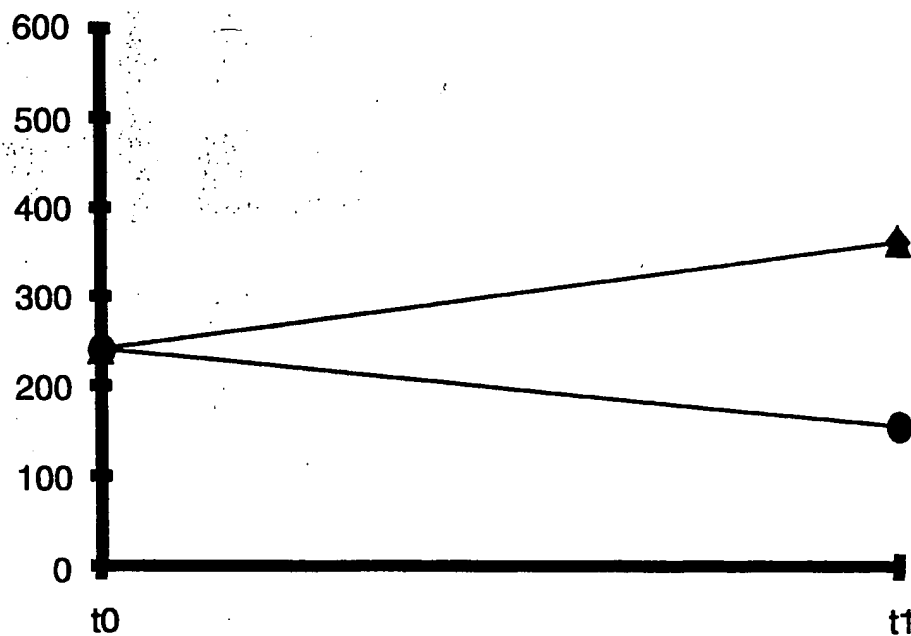


FIG. 4A



2/11

**FIG. 1C****FIG. 1D**

3/11

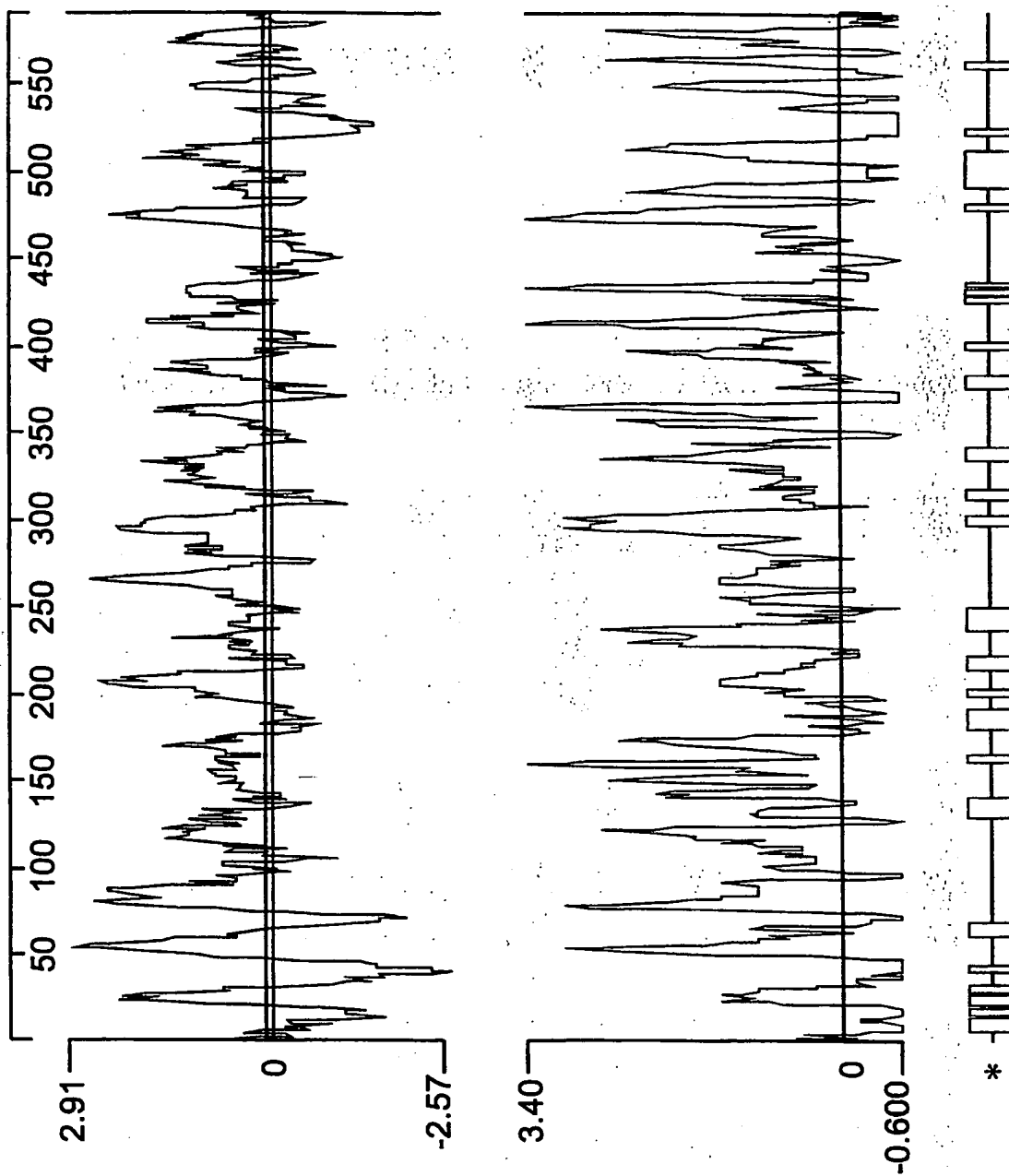


FIG. 1E

4/11

FIG. 2A

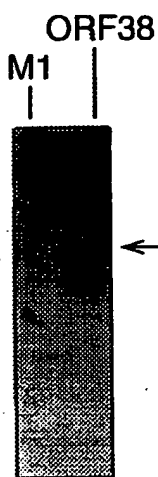


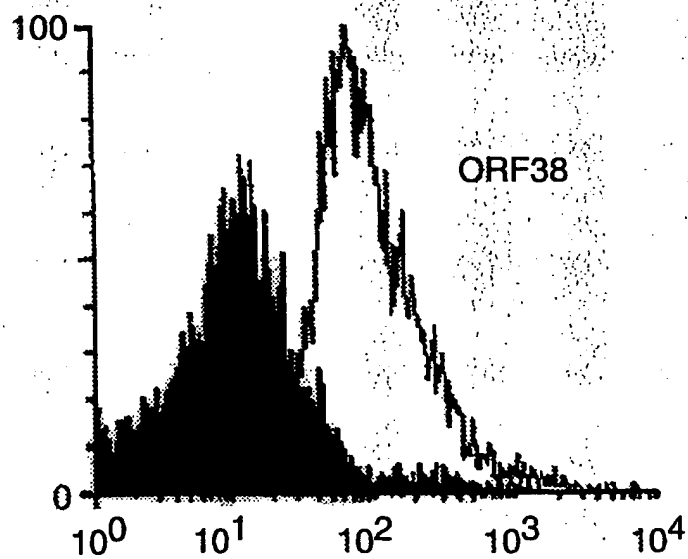
FIG. 2B



FIG. 2C



FIG. 2D



5/11

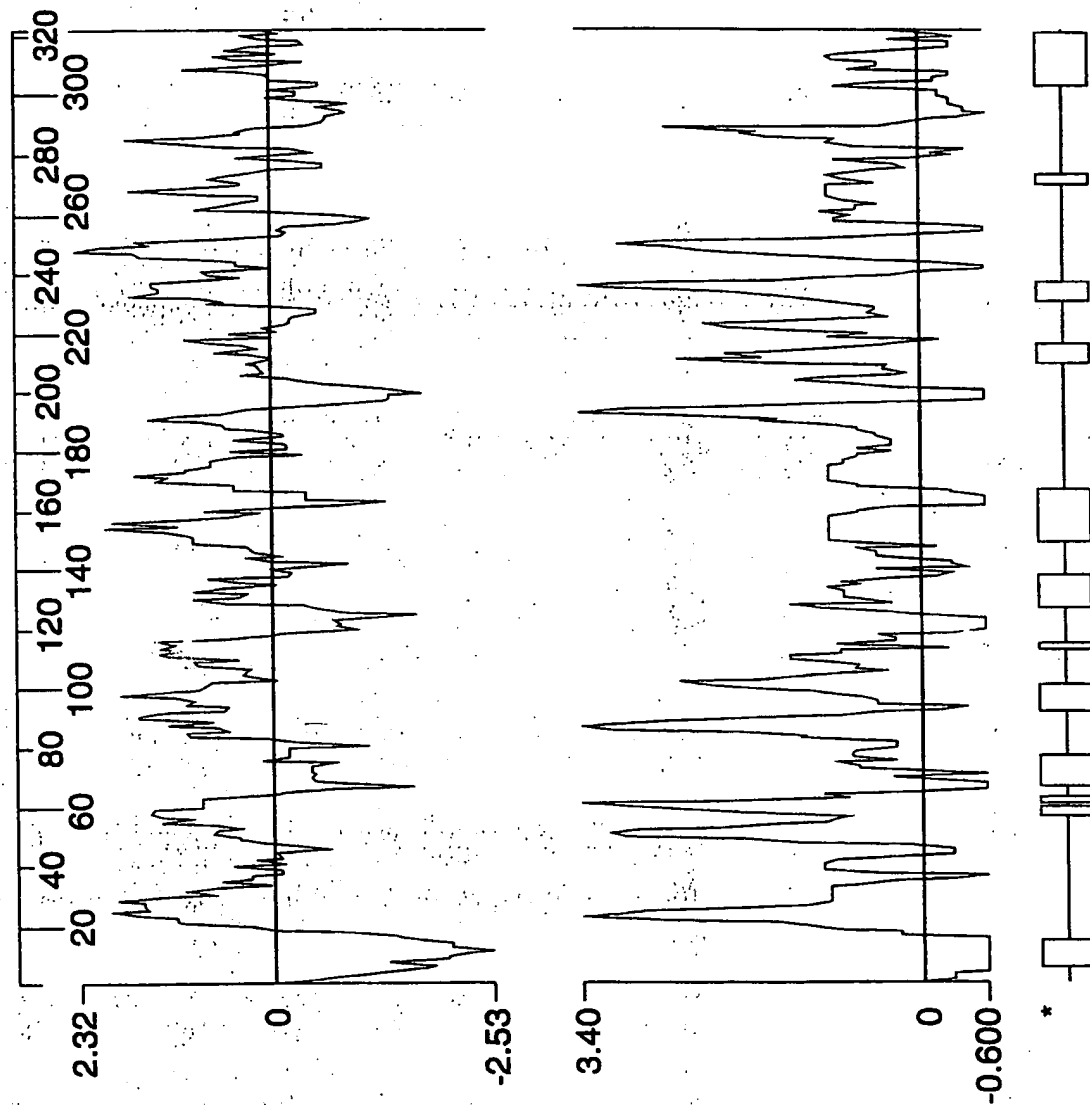


FIG. 2E

6/11

FIG. 3A



FIG. 3B

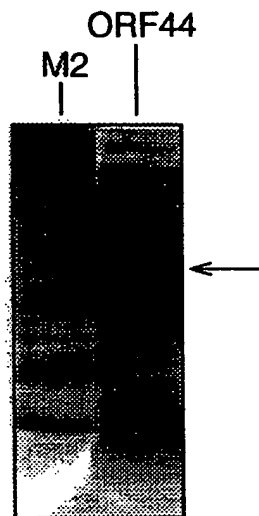
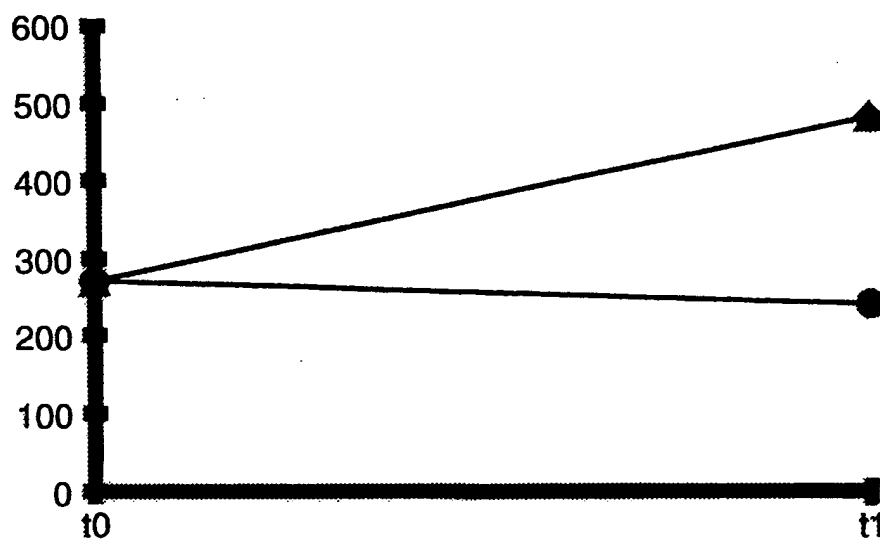


FIG. 3C



7/11

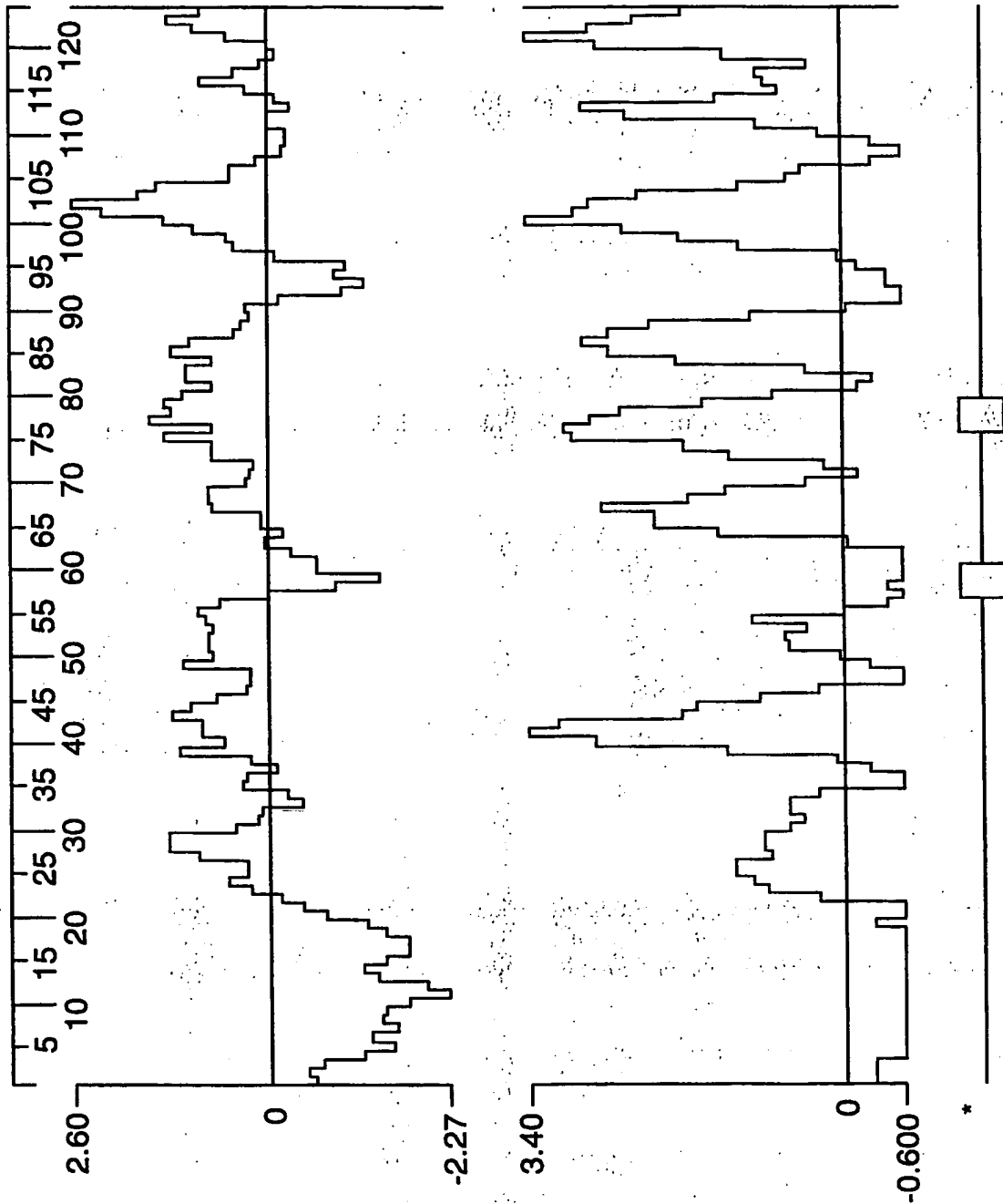
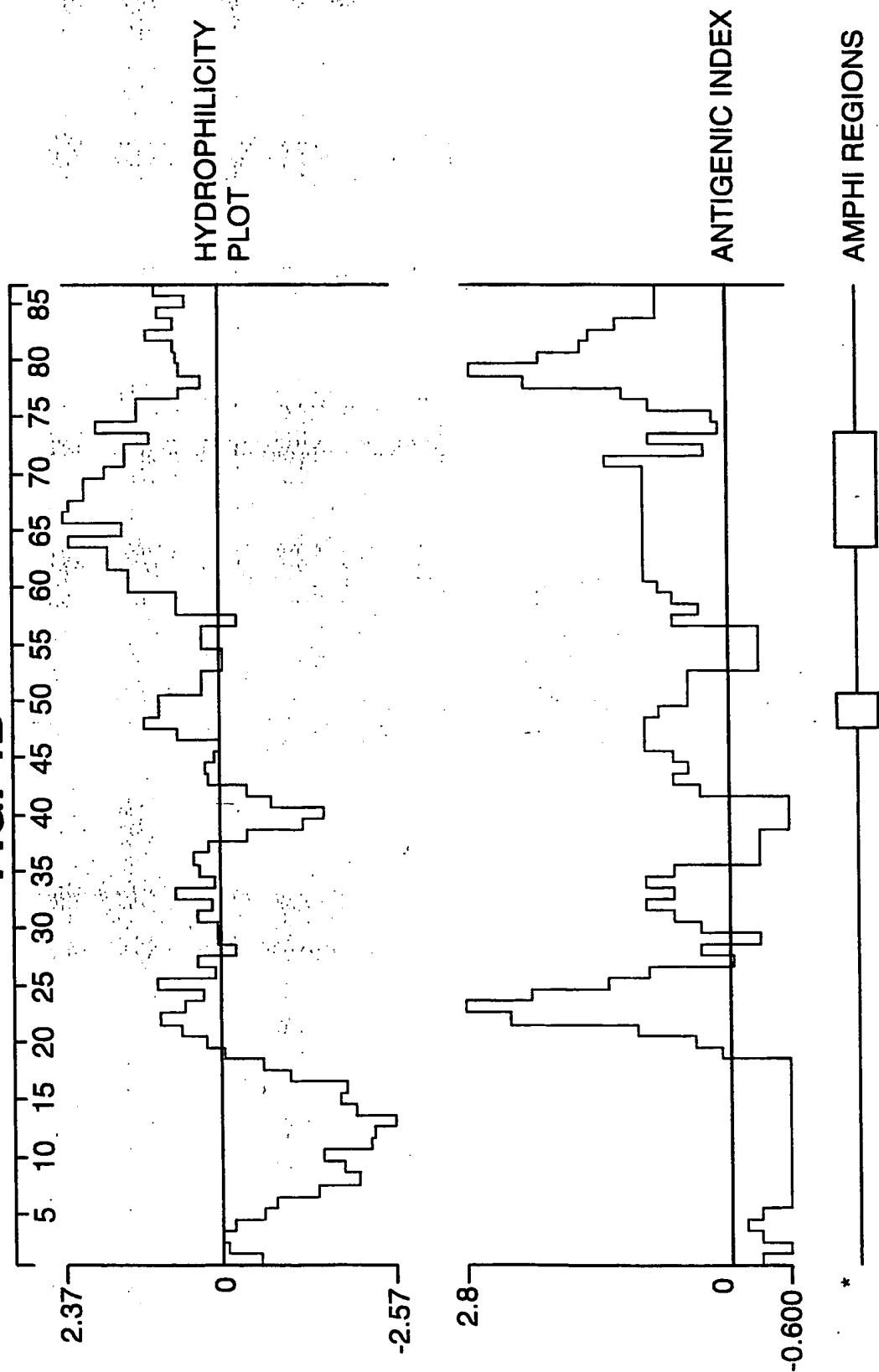


FIG. 3D

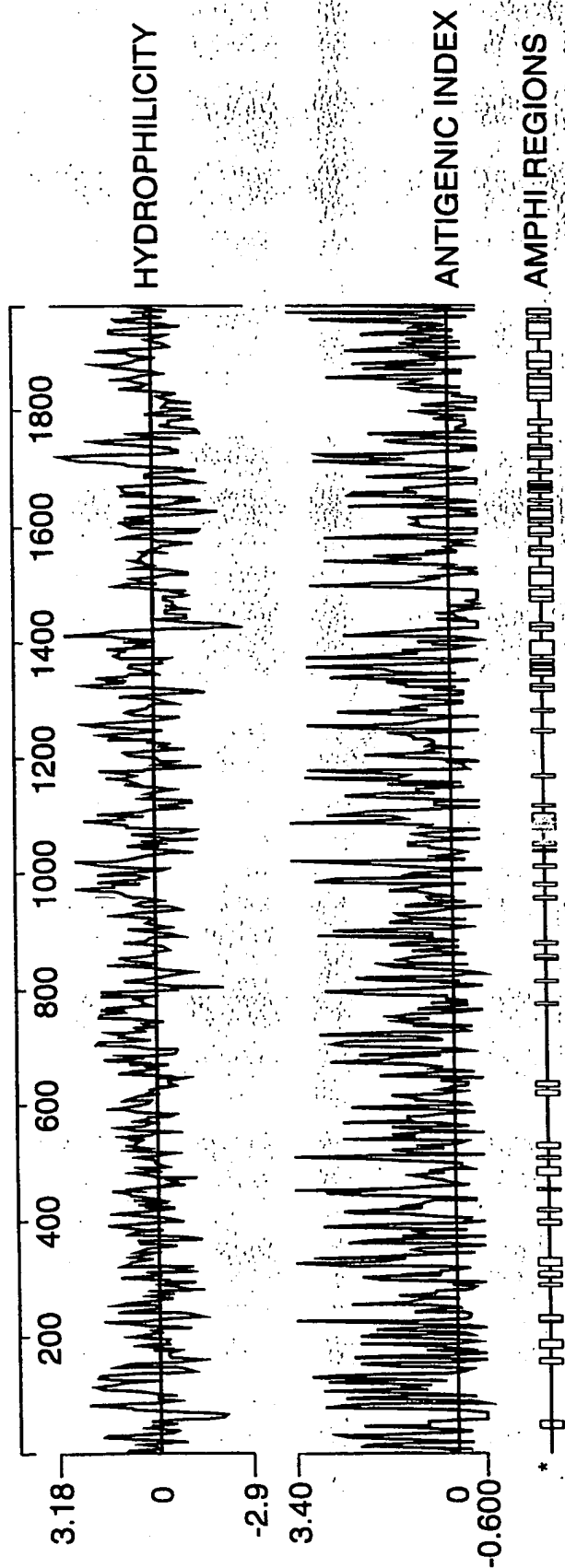
8/11

FIG. 4B



9/11

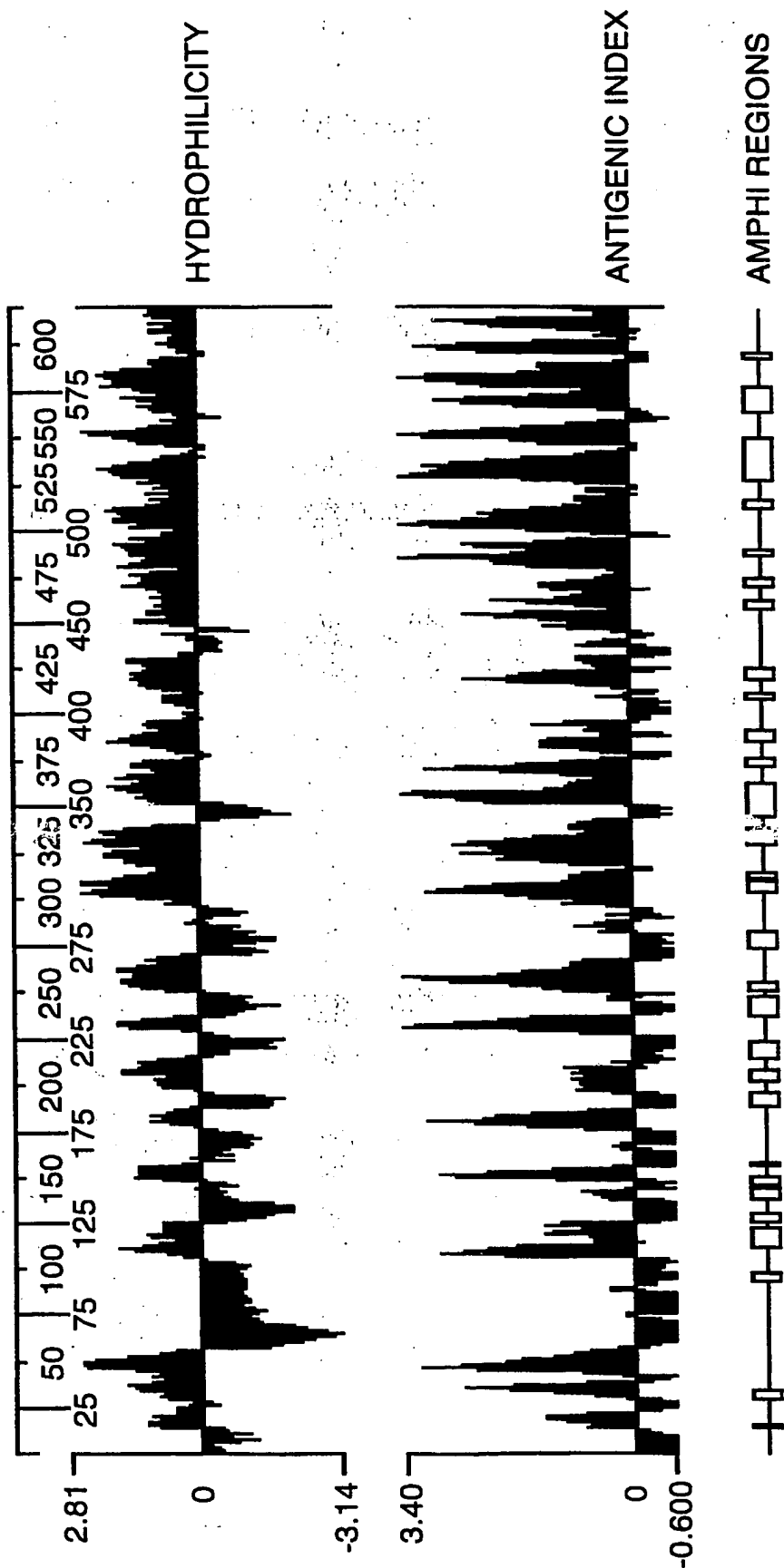
FIG. 5



Best Available Copy

10/11

FIG. 6



SUBSTITUTE SHEET (RULE 26)

Best Available Copy

11/11

FIG. 7

